

Clubbe, Richard (2018) *Serine integrase-based “landing pad” systems for chromosomal integrations of heterologous genes in Escherichia coli.*

PhD thesis.

<https://theses.gla.ac.uk/30585/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

**Serine Integrase-Based “Landing Pad” Systems For
Chromosomal Integrations of Heterologous Genes
in *Escherichia coli***

Richard Clubbe

M.Sc, B.Sc (Hons)

**Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy**

**University of Glasgow
College of Medical, Veterinary and Life Sciences
Institute of Molecular, Cell and Systems Biology**

2017

Abstract

Due to its long history of industrial applications, extensively studied biology and rapid growth rates, *Escherichia coli* is an important cell factory for metabolic engineering and an increasingly popular chassis organism for synthetic biology. Serine integrases, which function naturally to incorporate bacteriophage DNA into the host genome, represent a powerful and flexible tool for the genomic integration of heterologous genes or pathways into host organisms.

In the system described here, serine integrase enzymes are used to recombine two pairs of non-identical attachment sites, located on a chromosomally-integrated “landing pad” (LP) and an introduced “donor cassette”, resulting in cassette exchange of a transgene into a known locus on the *E. coli* genome. In this work, a landing pad was designed for ϕ C31 integrase, consisting of two ϕ C31 *attB* sites, both with differing central dinucleotide pairs, in a head-to-head orientation flanking an erythromycin resistance marker gene. This ϕ C31 LP was inserted into the *E. coli* genome, replacing the non-essential *pepA* gene, by recombineering. Cassette exchange experiments were then performed, in which a temperature-sensitive plasmid carrying a donor cassette consisting of corresponding *attP* sites to the ϕ C31 LP *attB* sites flanking an antibiotic resistance marker was transformed under the inducible expression of ϕ C31 integrase.

Serine integrases recombine their target sites in a highly unidirectional reaction. The reverse excision reaction, between integration product sites *attL* and *attR*, can only proceed with the addition of the integrase-associated recombinational directionality factor (RDF). Using *in vitro* recombination assays, TG1 integrase was shown to have very highly controlled directionality. Addition of the TG1 RDF gp25 strongly activates *attL-attR* recombination, taking recombination activity from a background level of 9% to 72% recombination of the substrate. TG1 gp25 also strongly inhibits *attP-attB* recombination, decreasing recombination from 77% to 8%. This demonstrates its potential as a highly controllable recombinase for synthetic biology. This work also investigated cross-reactivity of integrases and their non-cognate RDFs, showing that the TG1 RDF gp25 is able to activate ϕ C31 integrase *attL-attR* recombination, however does not similarly inhibit ϕ C31 *attP-attB*.

Abbreviations

<i>ApR</i>	Ampicillin resistance
<i>att</i>	Integrase attachment site
bp	Base pair
<i>CmR</i>	Chloramphenicol resistance
CRISPR	Clustered regularly interspaced short palindromic repeats
ddH ₂ O	Double distilled water
DICE	Dual integrase cassette exchange
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
<i>EmR</i>	Erythromycin resistance
FLP	Flippase
<i>Galk</i>	Galactokinase gene
GFP	Green fluorescent protein
<i>KnR</i>	Kanamycin resistance
LB	Lysogeny broth (media)
LP	Landing pad
PCR	Polymerase chain reaction
RDF	Recombinase directionality factor
RMCE	Recombinase mediated cassette exchange
SD	Synthetic defined (yeast media)
SDS	Sodium dodecyl sulphate
SIRA	Serine integrase recombinational assembly
ssDNA	Single stranded DNA
TAE	Tris acetate EDTA (electrophoresis buffer)
Tris	Tris(hydroxymethyl)aminimethane
TS	Temperature-sensitive
YKO	Yeast knockout
YPD	Yeast extract, peptone, dextrose (media)

Table of contents

Abstract.....	2
Abbreviations	3
Author's Declaration	8
Acknowledgements	9
1. Chapter 1: Introduction	10
1.1. <i>Escherichia coli</i> as a microbial cell factory.....	10
1.2. Chromosomal integration of heterologous genes.....	11
1.2.1. Advantages over plasmid expression.....	11
1.2.2. Currently available methods for chromosomal integrations of heterologous DNA	13
1.2.2.1. λ Red recombineering (homologous recombination).....	13
1.2.2.2. Phage recombinase-assisted integrations.....	15
1.2.2.3. CRISPR-Cas9.....	16
1.2.3. Limitations and challenges.....	18
1.3. Serine integrases as tools for synthetic biology.....	19
1.3.1. Serine integrases.....	19
1.3.1.1. Structure of the serine integrases	21
1.3.1.2. Serine integrase recombination mechanism.....	25
1.3.1.3. Recombination directionality factors (RDFs).....	27
1.3.1.4. Attachment site sequence requirements.....	29
1.3.1.5. Potential applications in synthetic biology.....	31
1.3.2. Current uses.....	34
1.3.2.1. Serine integrase recombinational assembly (SIRA).....	34
1.3.2.2. Serine integrase-mediated genomic integrations.....	34
1.4. A serine integrase-based genomic landing pad system.....	35
1.4.1. Landing pad system	35
1.4.2. Potential for highly efficient markerless genomic integrations.....	38
1.4.3. Towards a multi-integrase toolkit for whole pathway integration and strain optimisation.....	39
1.5. Challenges for developing an effective multi-integrase landing pad system.....	41

1.5.1. Determining the effect of genome position on landing pad integrations.....	41
1.5.2. Methods for target gene delivery to the landing pad.....	42
1.5.3. Orthogonality of serine integrases and their RDFs.....	43
1.6. Project aims.....	45
2. Chapter 2: Materials and Methods.....	48
2.1. Bacterial strains.....	48
2.2. Growth media.....	49
2.3. Antibiotics.....	50
2.4. PCR Primers.....	51
2.5. Plasmids.....	54
2.6. Restriction endonuclease digests.....	60
2.7. Ligation of plasmid DNA.....	60
2.8. Ethanol precipitation.....	60
2.9. Standard transformations of <i>E. coli</i> cells.....	61
2.9.1. Chemical transformation.....	61
2.9.2. Electrotransformation.....	62
2.10. Agarose gel electrophoresis.....	63
2.11. Extraction of DNA from agarose gel.....	64
2.12. DNA sequencing.....	64
2.13. PCR reactions.....	64
2.13.1. NEB Phusion polymerase.....	64
2.13.2. Taq polymerase.....	65
2.14. Purification of PCR product.....	65
2.15. λ Red recombineering.....	66
2.16. Curing cells of temperature-sensitive plasmids.....	67
2.17. Landing pad cassette exchange transformations.....	67
2.18. ISY100 random transposition of LP in <i>E. coli</i>	68
2.19. Screening of random LP mutants by arbitrary PCR.....	68
2.20. <i>in vitro</i> integrase reactions.....	69
2.21. <i>in vivo</i> <i>Galk</i> recombination assay.....	70
2.22. Yeast lithium acetate transformations.....	71
2.23. DNA quantification from agarose gels (digital).....	72

3.	Chapter 3: Demonstrating an integrase ‘landing pad’ (LP) in <i>E. coli</i>	73
3.1.	Introduction.....	73
3.2.	Design and construction of a ϕ C31 integrase landing pad.....	75
3.2.1.	Design of the LP construct.....	75
3.2.2.	Insertion of the LP cassette into the <i>E. coli</i> gene <i>pepA</i> by λ red recombineering.....	77
3.2.3.	Confirmation of LP insertion by PCR.....	81
3.3.	Development of an appropriate method for cassette delivery.....	85
3.3.1.	Designing temperature-sensitive donor vector.....	85
3.3.2.	Development of the ϕ C31 LP cassette exchange protocol (temperature-sensitive donor plasmid).....	88
3.4.	Addressing problems with ϕ C31 integrase expression in the LP strain.....	90
3.4.1.	Measuring ϕ C31 integrase activity by the GalK recombination assay.....	90
3.4.2.	Determining fate of ϕ C31 integrase expression plasmid in LPC311a cells.....	95
3.4.3.	An inducible ϕ C31 integrase expression method for improved activity in LPC311a.....	100
3.5.	Optimisation of ϕ C31 integrase LP cassette exchange protocol.....	103
3.5.1.	Transformation of donor cassette delivery vector into LPC311 cells with induced ϕ C31 integrase expression.....	103
3.5.2.	Detection of cassette exchange.....	104
3.5.2.1.	Antibiotic-resistance markers.....	104
3.5.2.2.	PCR of landing pad region.....	106
3.5.3.	Sequencing the ϕ C31 integrase LP on the LPC311a genome to investigate incomplete cassette exchange.....	114
3.6.	Discussion.....	116
4.	Chapter 4: Random transposition of several ϕ C31 LPs in <i>E. coli</i>	119
4.1.	Introduction.....	119
4.2.	Design of ϕ C31 LP ISY100 transposition cassette.....	120
4.3.	Determining the genomic locations of six randomly transposed ϕ C31 LPs by arbitrary PCR.....	122
4.4.	Confirmation of genomic LP locations by location-specific PCR.....	128

4.5.	Designing and cloning a temperature-sensitive donor cassette plasmid compatible with <i>KmR</i>	129
4.6.	Discussion.....	130
5.	Chapter 5: Investigating compatibility of integrases for a multi-LP system.....	132
5.1.	Introduction.....	132
5.2.	Measuring cross-compatibility of ϕ C31 and TG1 integrases and their corresponding RDFs.....	133
5.2.1.	ϕ C31 and TG1 integrase attL x attR recombination activation by non-native RDFs	133
5.2.2.	Inhibition of ϕ C31 and TG1 integrase <i>attP</i> x <i>attB</i> recombination by ϕ C31 gp3 and TG1 gp25 RDFs.....	139
5.3.	Discussion.....	143
6.	Chapter 6: Demonstrating a landing pad in <i>Saccharomyces cerevisiae</i> ...	147
6.1.	Introduction.....	147
6.2.	Establishing a ϕ C31 LP at several loci on the <i>S. cerevisiae</i> genome.....	148
6.2.1.	Designing a <i>S. cerevisiae</i> ϕ C31 LP.....	148
6.2.2.	Selecting genomic locations for LP integration.....	149
6.2.3.	Inserting the ϕ C31 LP into <i>KanMX</i> marker library strains by homologous recombination.....	151
6.2.3.1.	Cloning of ϕ C31 LP <i>KanMX</i> knockout delivery plasmid...	151
6.2.3.2.	Replacing <i>KanMX</i> with a <i>ura3</i> ϕ C31 LP by homologous recombination.....	155
6.3.	Discussion.....	159
7.	Chapter 7: Discussion.....	161
7.1.	Establishing a ϕ C31 integrase landing pad system in <i>E. coli</i>	161
7.2.	Non-cognate activity of integrases and other RDFs.....	163
8.	References.....	165

Author's Declaration

The research reported in this thesis is my own work unless otherwise stated. The work has not been submitted for any other purpose.

Richard Morton Clubbe

Acknowledgements

Firstly I would like to thank my supervisor, Professor Marshall Stark, whose support and patience were vital throughout this PhD. Special thanks also go to Dr Femi Olorunniji for the time and effort he gave to helping me in the lab and the directional advice he provided for the project as a whole. The Hooker lab has been a fantastic place to work and I particularly appreciate the help of Arlene McPherson whose hard work makes it what it is. Thank you also to Dr Sean Colloms, whose work was the inspiration and basis for much of this project, and to my second supervisor Dr Joseph Gray for his timely advice. Significant credit for this PhD go to my parents Robert and Jill for their incredible support, and to my friends who did a perfect job of making the harder times bearable! Final thanks go to the BBSRC for funding this work and the WestBio DTP for giving me the opportunity.

1. Chapter 1: Introduction

1.1 *Escherichia coli* as a microbial cell factory

Escherichia coli has a long and successful history as a microbial cell factory for expressing commercially-relevant proteins on an industrial scale. Even before the advent of recombinant DNA technologies it was one of the first microorganisms in which a protein was produced by over-expression, with the cellular yield of lactose repressor increased through selection and phage-mediated expression of an high-expression *lacI* mutant (Muller-Hill *et al.*, 1968). After this, the emergence of recombinant DNA techniques such as the use of restriction enzymes and *in vitro* DNA recombination introduced the ability to express non-native, heterologous genes in *E. coli* cells. The first biotechnological drugs produced using these techniques included insulin, growth hormone and interferons. Between 1982 and 2013, there have been 23 drugs produced in *E. coli* have been approved for use as pharmaceuticals (Baeshen *et al.*, 2015). In fact, 30% of the approved therapeutic proteins being produced today use *E. coli* as an expression host.

As a highly capable host for producing stably folded, globular proteins from both prokaryotic and eukaryotic organisms, there are many advantages to choosing *E. coli* for expressing recombinant proteins. It has an unrivalled growth rate among industrially-relevant microorganisms, with a doubling time of around 20 minutes, meaning that it only takes a few hours for a culture to reach stationary phase (Sezonov *et al.*, 2007) with an upper potential cell density of around 1×10^{13} cells per mL (Shiloach & Fass, 2005). On top of this, *E. coli* cells can also achieve very high protein synthesis rates (Sahdev *et al.*, 2008), in some cases reaching expression levels as high as 30% of the total cellular protein (Baeshen *et al.*, 2015). *E. coli* culture media is also inexpensive and can be made using easily available ingredients, making mass bioculture relatively cheap for industrial purposes.

When designing a strategy for recombinant protein production, there are many strains of *E. coli* available to choose from, with each tailored for a specific function or production requirement, as well as basic expression strains that provide a simple background for rapid first expression screening (Rosano & Ceccarelli, 2014). In its years of use as an heterologous host, extensive libraries of expression vector systems have accumulated, with a choice of plasmid parts that allow customised copy number, promoter type, selective markers and tags for detection or purification. However, due to the precision required for synthetic biology cell engineering systems, many of the shortcomings of plasmid-based expression mean that chromosomal integration of heterologous genes is preferable.

1.2 Chromosomal integration of heterologous genes

1.2.1. Advantages over plasmid expression

The traditional method for heterologous gene expression in *E. coli* and other hosts is via plasmid vector expression. This method, for which there are numerous libraries of expression vectors, will always be useful for expressing target genes on a small scale, for example for purifying a protein of interest in the lab. However, as scale of expression and microbial culture increases, the plasmid vector expression system begins to become restrictive. Population dynamics and growth states tend to have an effect on plasmid copy numbers and gene expression levels, making it hard to control expression levels in industrial cultures when using plasmid vectors (Lin-Chao & Bremer, 1986). Plasmid propagation and maintenance places a burden on cells which may already be under significant metabolic strain to express the target protein (Birnbaum & Bailey, 1991).

In addition, to be maintained in cells many plasmid vectors require constant selective pressure, often by antibiotic resistance, which places further strain on cells and in larger scale operations can become expensive. Problems with plasmid instability can also be encountered, particularly in the case of products which may be harmful to the cells such as toxic biopharmaceutical and biofuel products. In addition to potential issues during large scale culture, assembly and optimisation of plasmid vectors comprising several genes is a time-consuming process in which several cloning and screening steps are necessary. This can prove a rate-limiting step during metabolic engineering and strain development. In recent years there has been a greater focus on the integration of heterologous genes into the microbial host genome. This is desirable, firstly as it overcomes problems inherent with plasmid vector expression, removing the need for constant selective pressure, reducing the metabolic burden placed on the cell by plasmid maintenance and avoiding problems of instability. Secondly, genomic integration of a target gene can allow for a higher level of regulatory control, removing expression fluctuations often seen in plasmid expressions.

1.2.2. Currently available methods for chromosomal integrations of heterologous DNA

1.2.2.1. λ Red recombineering (homologous recombination)

One of the most common approaches to genomic integration of heterologous genes in *E. coli* is a method of λ phage-assisted homologous recombination, termed 'recombineering' (Copeland *et al.*, 2001). Homologous recombination is a natural cell mechanism performed by *E. coli* to fix incompletely replicated chromosomal DNA during replication and to repair single- and double-stranded damage to chromosomal DNA. In *E. coli* the genes that encode this machinery are the Rec genes (Kowalczykowski *et al.*, 1994). The process repairs a stretch of DNA by seeking out another copy of the same DNA sequence in the cell, identified by sequence homology, unzipping the broken DNA and recombining the two DNA molecules to produce two identical in-tact double stranded DNA sites.

In recombineering, the homologous recombination mechanism is hijacked by three λ Red proteins: Gam, a nuclease inhibitor, which prevents the *E. coli* nuclease RecBCD from binding and unwinding the linear DNA to be integrated (that is introduced to the cell via transformation); Beta, a single strand DNA binding protein that aids annealing of homologous sequences on the target site and introduced linear DNA; and Exo, a 5' to 3' exonuclease, which creates that creates the single strand overhangs to which Beta binds (Sharan *et al.*, 2009). The recombineering approach requires that the DNA being introduced be amplified by PCR with primers that attach homologous sequences (around 50 bp) at each end. This PCR product is transformed into cells under expression of the λ Red elements in order to protect the linear PCR product and promote homologous recombination. A diagram of the recombineering process is shown in Figure 1.1.

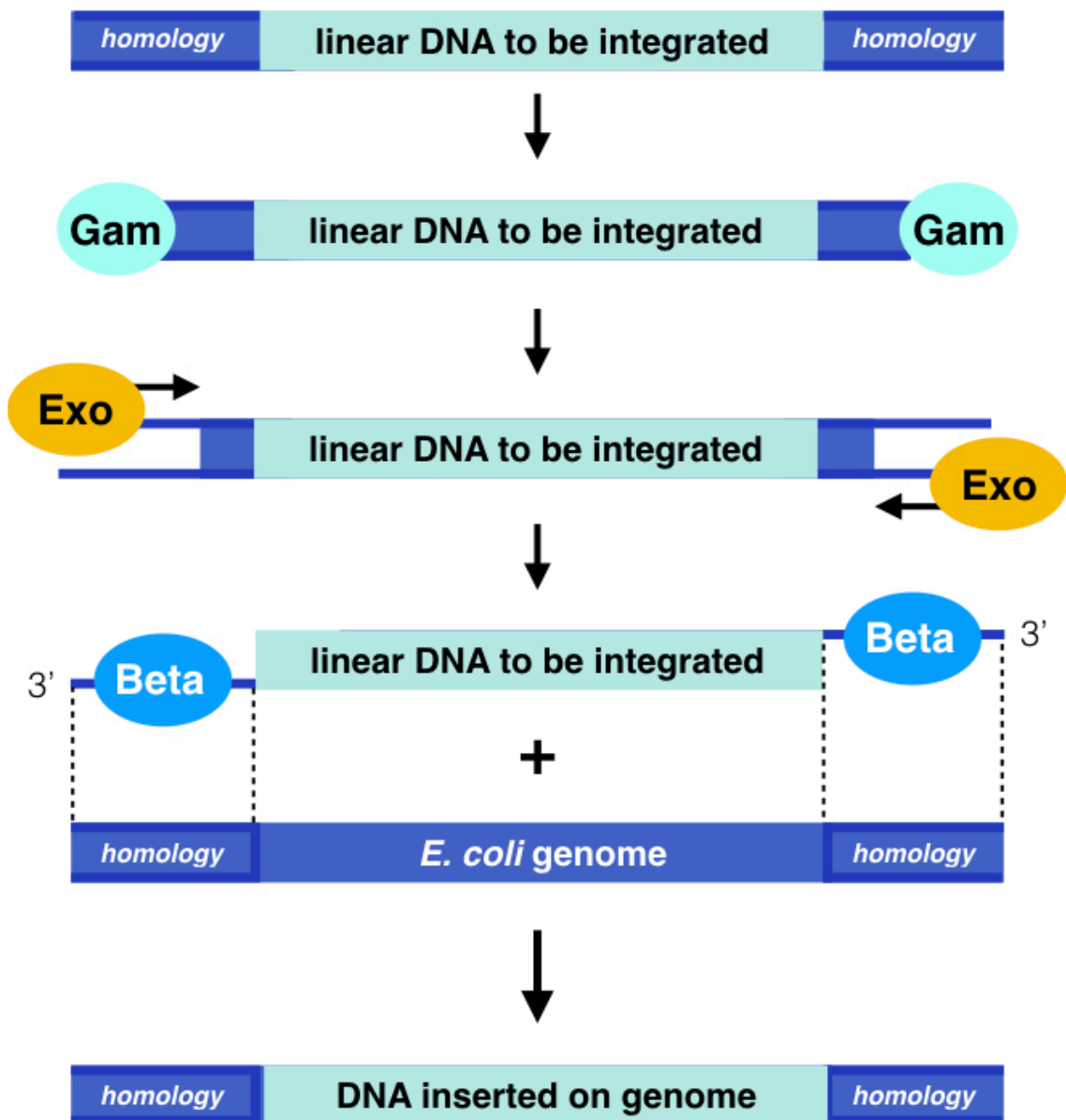


Figure 1.1: Overview of the λ Red recombineering method. The gene to be inserted into the genome amplified by PCR to be flanked by sequences homologous to the site on the genome being targeted. Gam first protects the inserted linear DNA from degradation by *E. coli* nucleases. The λ 5' to 3' exonuclease Exo creates 3' overhangs on the linear DNA homology region. Beta then binds to the single stranded overhangs and promotes annealing and homologous recombination with the genomic target site, resulting in the gene being inserted onto to *E. coli* genome at the predetermined site. Based on Sharan *et al.*, 2009.

1.2.2.2. Phage recombinase-assisted integration

The earliest examples of site-specific recombinases being utilised for genomic rearrangements in prokaryotes or eukaryotes were the tyrosine recombinase Cre, which originates in bacteriophage P1 and targets short inverted sites called *loxP*, and FLP (flippase), taken from the 2- μ circle plasmid of *Saccharomyces cerevisiae* which targets *FRT* recognition sequence sites. These have been used widely for genomic integrations and rearrangements in bacteria and yeast. The advantage of these recombinases are that they are small in size, target short attachment sequences and require no cofactors for effective recombination. However, since they target and recombine two identical sites (*loxP* x *loxP* or *FRT* x *FRT*), these enzymes catalyse the reverse DNA excision reaction as well as the integration reaction. This is due to the fact that the recombinant product sites have the same sequence as the target substrate sites, and means that genomic integrations using these recombinases don't always show the high efficiencies demanded in synthetic biology.

The Cre-*loxP* system has been used extensively for gene deletions, rearrangements and heterologous gene integrations, mostly in eukaryotic cells (Sauer, 1998). Another use of the Cre-*loxP* system has been the the secondary removal of marker genes or plasmid backbone after integration into the genome (Martinez-Morales, 1999). However, most recombinase-mediated genomic integration methods in *E. coli* have involved the integration of whole plasmids into naturally-occurring target sites on the genome. Several other phage recombinases have been demonstrated to do this in *E. coli*, including the λ , HK022, ϕ 80, P21, and P22 phages (Haldimann & Wanner, 2001).

The use of recombinases for genomic integrations in *E. coli* has also been developed into a recombinase-mediated cassette exchange (RMCE) method, which replaces a target cassette on the chromosome flanked by two non-compatible, or 'heterospecific' attachment sites, with a plasmid-borne cassette featuring two corresponding attachment sites. This was first done with FLP recombinase and heterospecific *FRT* target sites in mammalian cells

(Shlake & Bode, 1994). However, this method relies on the incompatibility of the two non-identical *FRT* sites variants, which can sometimes recombine and cause loss of the target cassettes from the genome or the delivery plasmid once FLP recombinase is introduced. The RMCE method has been demonstrated for both FLP and Cre (Turan *et al.*, 2011).

1.2.2.3. CRISPR-Cas9

In recent years, CRISPR-Cas9 has become an important tool for genome editing, particularly in eukaryote cells. The system makes use of the Cas9 DNA endonuclease derived from CRISPR (clustered regularly interspaced short palindromic repeats) bacterial immune systems. The Cas9 endonuclease cuts at sequences specified by its associated guide RNA, creating a double strand break that must be repaired by native cell machinery. Naturally, the Cas protein is guided by two CRISPR RNAs, however in seminal work by Jinek *et al* (2012) these RNAs were fused to create a single guide RNA (sgRNA) that simplified the system for use in genome engineering. The introduced break can then be repaired by non-homologous end joining (NHEJ) to create mutations at the target site, or homology directed repair (HDR) which introduces donor DNA into the break site (Wang *et al.*, 2016). The power of Cas9 as a genome editing tool lies in the ability to synthesise sgRNA to specify the sequence at which it will cut, therefore making the system highly programmable and able to introduce mutations or new DNA at any site of choosing. For an excellent review of the wide applications and potential of CRISPR-Cas9 for genome editing in eukaryotes see Wang *et al*, 2016.

The precise and programmable nature of CRISPR-Cas9 have made it a revolutionary tool for genome editing. However, there are still issues with off-target effects due to the small size of the 20 bp sgRNA which can show homology with other sequences present on the genome, as well as other problems with efficiency and delivery (these shortcomings are reviewed by Peng *et al*, 2016). Most of the successful applications of CRISPR-Cas9 have been based on introducing point mutations into the genome by non-

homologous end joining, rather than the introduction of long stretches of new DNA, which is the objective of the serine integrase system in this thesis.

While most CRISPR-Cas9 work has focussed on gene editing in eukaryote cells, the system has also been developed for high-efficiency gene editing in *E. coli*. In 2013, it was demonstrated that the CRISPR-Cas9 system could be applied in *E. coli* and *Streptococcus pneumoniae* cells to introduce marker free mutations via homology directed repair (Jiang *et al.*, 2013). Since then CRISPR-Cas9 gene editing in *E. coli* has been further enhanced by coupling the mechanism with lambda recombinase-mediated homologous recombination (recombineering), creating a high-efficiency method for mutating multiple genes with accuracy, a method that was also first suggested by Jiang *et al* (2013).

Of particular relevance to the work in this thesis, the coupling of CRISPR-Cas9 to lambda recombinase-mediated homologous recombination (recombineering) in *E. coli* has led to development of methods for more effective integration of large stretches of heterologous DNA. In 2015, Pyne *et al.* demonstrated the ability of the CRISPR-recombineering method to integrate PCR products of up to 3 kb at efficiencies that allowed successful mutants to be identified by colony PCR. More recently, Bassalo *et al.* (2017) demonstrated that their CRISPR-recombineering method was able to integrate a large markerless pathway (for isobutanol production) of around 10 kb into different sites on the genome at an efficiency of around 50%.

CRISPR has also been used to improve efficiency of gene integrations in yeast. Introducing highly specific double strand breaks in *S. cerevisiae* has been shown to promote homologous recombination (in yeast a much more powerful native mechanism than that of *E. coli*) and therefore improves integration of introduced homology-flanked linear DNA (Storici *et al.*, 2003). Recently, the use of CRISPR has been shown to achieve very high efficiencies of gene integrations at several sites on the *S. cerevisiae* genome by combining it with the genome engineering tool EasyClone (Ronda *et al.*, 2015).

1.2.3. Limitations and challenges

The primary shortcomings of most genomic integration methods are the low integration rates achieved and propensity for off-site errors. These outcomes mean that there is a requirement for time-consuming screening processes following the integration protocol to ensure that the target gene has integrated correctly and only in the target location. The most commonly used genomic integration method used in *E. coli*, λ Red-mediated homologous recombination, in particular generally has low efficiencies, which are dependent on the size of DNA to be integrated and the length of the flanking homologous sequences (Wang *et al*, 2009; Mosberg *et al*, 2012). For all its promise, CRISPR has often been limited in its application in *E. coli* due to the organisms poor ability to repair double strand breaks, which results in lethality (Luo *et al.*, 2016), as well as its tendency off-target activity, low efficiencies for long DNA sequence integrations and issues with Cas9 activity (Peng *et al*, 2016).

Low efficiencies of genomic integrations from existing methods for the most part mean that a marker or selection gene must always be included in the insert, which is undesirable in some strain development contexts. The tools that overcome these problems, such as phage-mediated recombinations, are mostly limited to naturally-occurring attachment sites and have usually been used to integrate entire plasmids rather than single genes. Even these recombination methods are not completely efficient, as the enzyme that facilitates integration of the target DNA also catalyses the reverse excision in a bidirectional reaction, meaning that a very high concentration of donor plasmid is required for effective integration. Therefore there is still clearly a demand for a unidirectional genomic integration method that is capable of high efficiency integrations of markerless target DNA into a chosen location on the genome.

1.3 Serine integrases as tools for synthetic biology

1.3.1. Serine integrases

Serine integrases represent a potentially powerful and flexible tool for synthetic biology and the genomic integration of heterologous genes. They belong to the serine recombinases, one of two families of site-specific recombinases alongside the tyrosine recombinases (which includes Cre, FLP and λ), both named and organised according to the amino acid residue that bonds to the DNA during catalysis (Grindley *et al.*, 2006). The majority of known serine integrases belong to bacteriophages, which use these enzymes to incorporate bacteriophage DNA into its bacterial host genome. However, serine integrases have also been discovered to function in the transfer of antibiotic resistance genes via transposon (Wang & Mullany, 2000) or other genetic elements (Katayama *et al.*, 2000).

In their natural bacteriophage context, serine integrases bind to attachment sites *attB* and *attP*, on the bacterial and phage genomes respectively, catalysing the integration of the phage genome and producing two product sites, *attL* ('Left' of the integrated DNA) and *attR* ('Right' of the integrated DNA). These *attL* and *attR* sites consist of two half sites from the *attP* and *attB* sites (Figure 1.2). The integrase also catalyses excision of the prophage from the bacterial genome but only in the presence of the integrase's recombination directionality factor (RDF), a phage-encoded protein which facilitates recombination between integration product sites *attL* and *attR* in a reverse reaction. A diagram representing this integration/excision of phage DNA is shown in Figure 1.2.

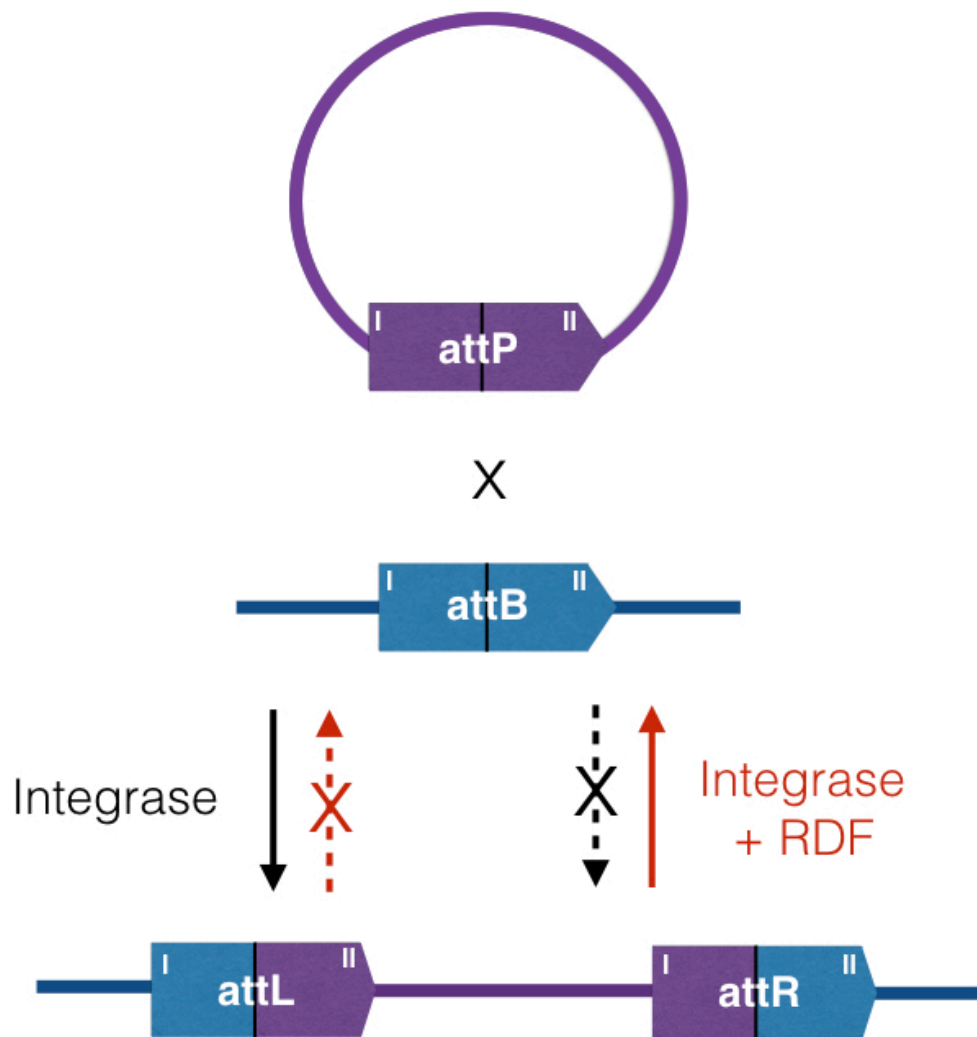


Figure 1.2: Simple diagram illustrating how serine integrases recombine an *attP* and *attB* site to integrate phage DNA into the bacterial chromosome to form product sites *attL* ('Left') and *attR* ('Right'). The *attL* product site consists of a half-site from *attB*, *B'* (blue), and a half site from *attP*, *P''* (purple), while *attR* is made up of the other two half sites, *P'* (purple) and *B''* (blue). The recombination reaction between these sites to excise the prophage can only occur in the presence of the serine integrase's recombination directionality factor (RDF).

1.3.1.1 Structure of the serine integrases

Serine integrases, such as the *Streptomyces* bacteriophage ϕ C31 integrase of focus in much of this work, generally consist of a catalytic N-terminal region of around 130 to 150 amino acids and a much larger C-terminal region of between 300 and 500 amino acids, which is involved in the targeting and binding of attachment site sequences (Smith and Thorpe, 2002). These two regions are connected by a long helix (α E) which is significantly involved in dimerisation, tetramerisation and DNA-binding during recombination (Rutherford & Van Duyne, 2014).

While initial studies on serine integrases focussed on the integration activity and biochemical characteristics of the enzymes, actual structural and mechanistic details of their activity weren't established until more recently. The structure of the C-terminal region was first revealed in a 2013 study by Rutherford *et al*, which examined the structure of the *Listeria innocua* prophage LI integrase bound to its *attP* half-site. This C-terminal region, which is involved in most of the integrase's DNA-binding activity, can be considered to comprise two DNA-binding domains: the 'recombinase domain' and the zinc ribbon domain (Van Duyne and Rutherford, 2013). Within the zinc ribbon domain is an extended coiled-coil motif which has important function in recombination directionality (described below). Figure 1.3 shows how these structures are defined in six serine integrases according to alignment with A118 integrase, the crystal structure of which was identified as an integrase-*attP* half-site complex by Rutherford *et al*. (2013).

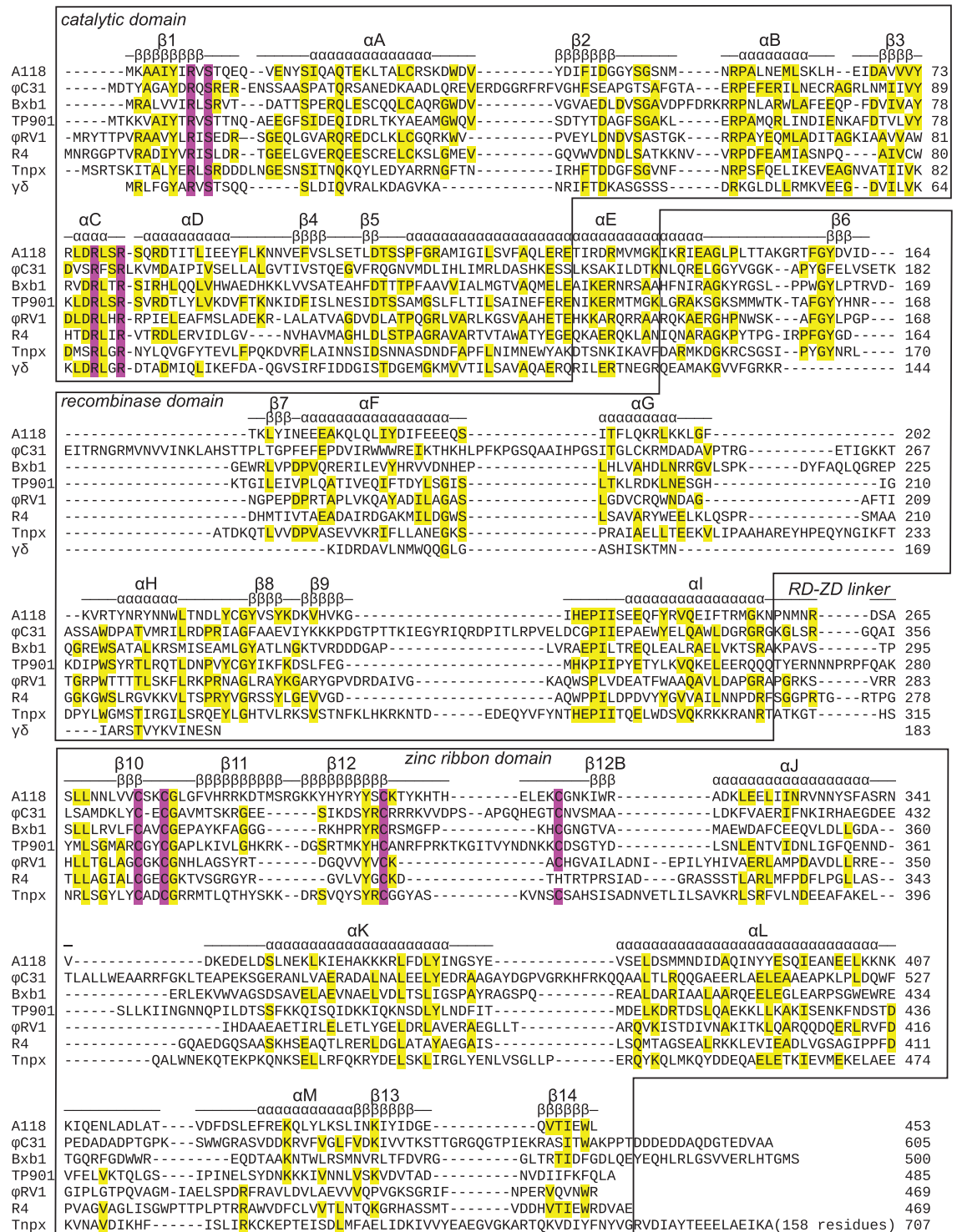


Figure 1.3: Alignment of protein sequences of seven serine integrases: A118, φC31, Bxb1, TP901-1, φRV1, R4 and TnpX. Secondary structures labelled here are based on the above-mentioned structural studies on the TP901-1 C-terminal domain (Yuan *et al.*, 2008) and A118/attP half-site complex (Rutherford *et al.*, 2013) in conjunction with the A118 integrase sequence identified by Loessner *et al.* (2000). Residues that are conserved in three or more of these integrases are shaded, with conserved catalytic residues and zinc-coordinating cysteines in bold. Image: Van Duyne and Rutherford, 2013.

The structure of the serine integrase catalytic N-terminal domain has been found to be very similar to that of the better-understood resolvase and invertase serine recombinases, such as Sin resolvase (Mouw *et al.*, 2008). Evidence for this has come from elucidations of the structure of the tetrameric form of TP901-1 integrase (Yuan *et al.*, 2008), and two-domain structure of ϕ C31 integrase (Rutherford & Van Duyne, 2014).

To illustrate the general overall structure of the serine integrases, the tetrameric structure of TP901 integrase (as it is during synapsis) established by Yuan *et al.* is shown in Figure 1.4, while the dimeric structure of the TP901 C-terminal domain (as it is when it binds to the attachment site DNA before forming a tetramer for synapsis) is shown in Figure 1.5 (Van Duyne and Rutherford, 2013). These can be considered alongside the further structural and sequence information shown in Figure 1.3.

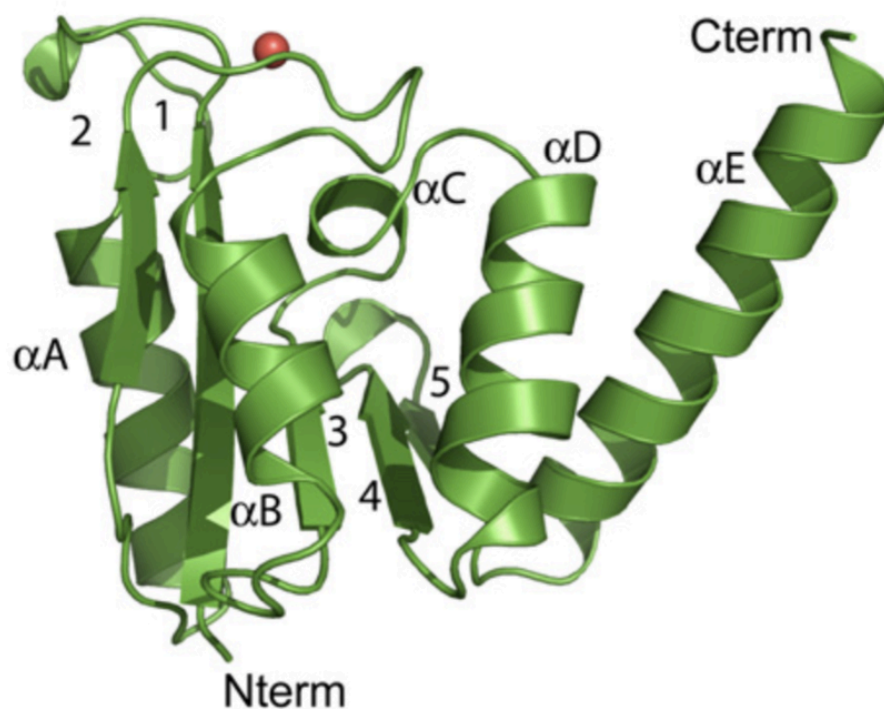


Figure 1.4: Tetrameric structure of the TP901 integrase. The catalytic C-terminal domain (Cterm) DNA-targeting N-terminal domain (Nterm) are labelled. See also protein sequence and structural annotation in Figure 1.3. Image: Yuan *et al.*, 2008.

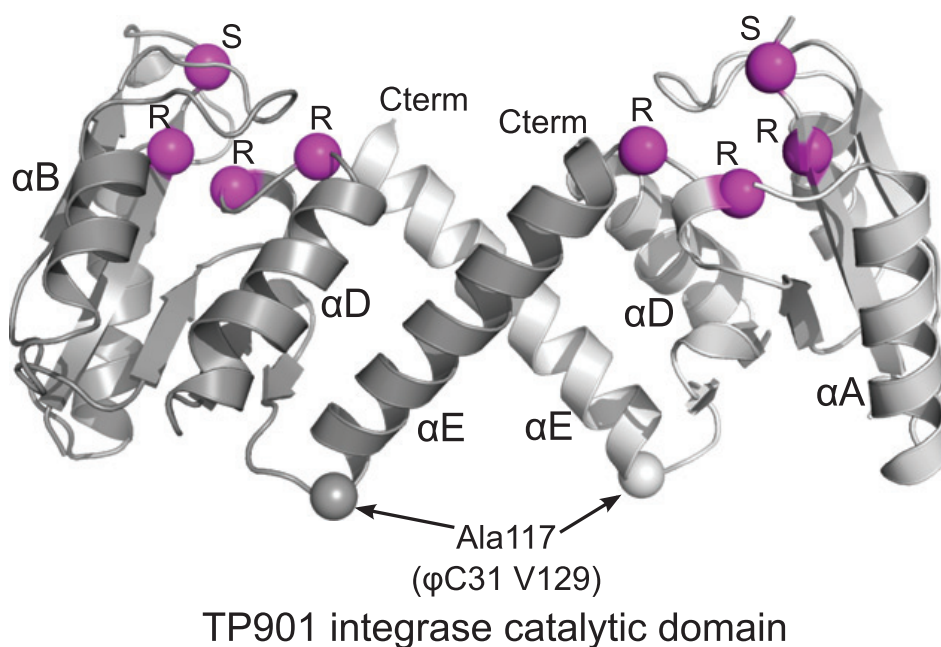


Figure 1.5: Structure of the TP901 catalytic C-terminal domain as a dimer. Catalytic residues are represented by magenta spheres. Also indicated is the TP901 Ala117 which corresponds in position to Val129 in ϕ C31 integrase (see Figure 1.3). Image: Van Duyne and Rutherford, 2013.

1.3.1.2 Serine integrase recombination mechanism

In their review, Rutherford and Van Duyne (2014) collected the to-date evidence available on the serine integrase functional mechanism to propose a widely referenced and accepted model of integration. This serine integrase recombination model is shown in Figure 1.6. Upon recognition of its attachment site sequence, the integrase forms a dimer with another integrase unit, with each subunit bound to a half-site of the attachment site. See Figure 1.5 for a structural illustration of TP901 integrase forming this dimer as it is when bound to its attachment site.

A pair of *attP* and *attB* sites bound by integrase dimers are then brought together to form a tetramer. This complex, consisting of an integrase tetramer bound to the two attachment sites, is referred to as a synaptic complex (Grindley et al, 2006). The structure of an integrase subunit of TP901 when in a tetrameric synaptic complex is shown in Figure 1.4. It is thought that the synaptic complex is stabilised by interaction of the zinc ribbon domain coiled-coil motifs of opposite integrase monomers (see Figure 1.6).

Within the synaptic complex, DNA cleavage occurs to leave a 2 bp overhang at the centre of each attachment site sequence. Cleavage is followed by a vertical 180° rotation of two half-site-bound integrase subunits at odds with the rest of the synaptic complex in order to connect each *attP* half-site with its corresponding *attB* half-site. After this rotation, ligation occurs between the opposite attachment site half-sites in order to form the two recombination product sites *attL* (Left) and *attR* (Right) (Smith, 2015).

The new *attL* and *attR* sites, still bound to integrase dimers, then dissociate from the synaptic complex. Following this, it is thought that the coiled-coil motifs within the zinc ribbon domain interact intra-molecularly in a way that prevents the reverse reaction between *attL* and *attR* from occurring, thus locking the dimer-*attL/attR* complexes and ensuring that the *attP-attB* reaction proceeds unidirectionally (Rutherford and Van Duyne, 2014).

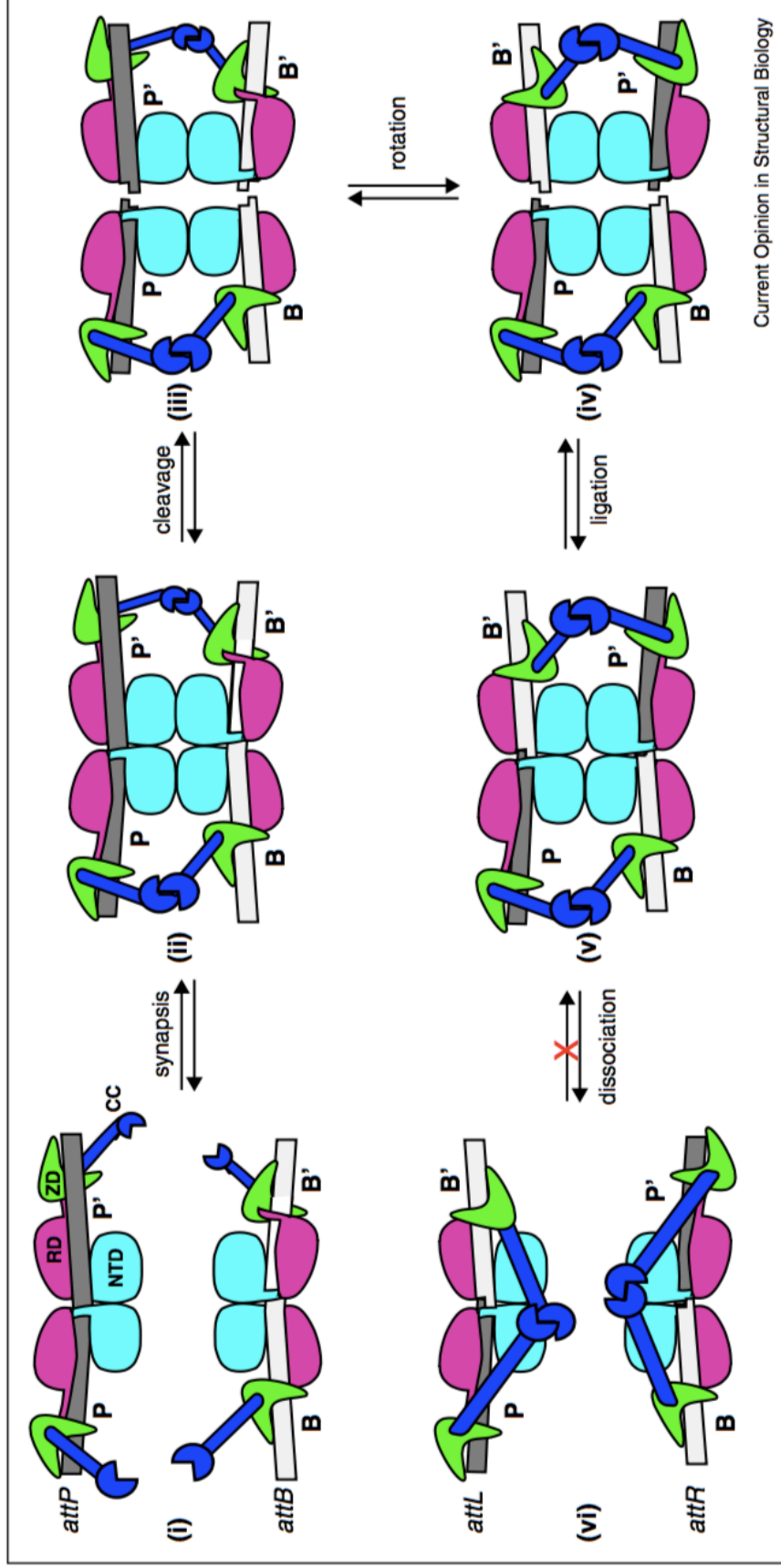


Figure 1.6: Current model for serine integrase recombination of *attP* and *attB* attachment sites. Integrase functional structures are labelled as follows: NTD, N-terminal catalytic domain (cyan); RD, recombinase domain (magenta); ZD, zinc ribbon domain (green); CC, coiled-coil motif (blue). (i) integrase dimers bind to *attP* and *attB* sites; (ii) DNA-bound integrase dimers form a tetrameric synaptic complex, stabilised by interaction between coiled-coil motifs of the zinc ribbon domain; (iii) each integrase subunit cleaves one of the four DNA strands within the synaptic complex, creating 3'-dinucleotide overhangs; (iv) integrase subunits bound to half-sites P' and B' rotate around 180°; (v) ligation occurs between complementary dinucleotide overhangs to create two new sites, *attL* (P & B') and *attR* (B & P'); (vi) integrase dimers dissociate from the synaptic complex bound to *attL* and *attR*, with interactions between coiled-coil motifs preventing consequent synapsis occurring for the reverse *attL-attR* recombination reaction. The structural conformation of the integrase when attached to an *attP* or *attB* half-site also described by Rutherford and Van Duyne (2014) is shown in Figure Y. Image: Rutherford and Van Duyne, 2014.

1.3.1.3 Recombinase directionality factors (RDFs)

For the reverse recombination reaction to occur between product sites *attL* and *attR*, serine integrases require the presence of their associated recombination directionality factor (RDF). RDFs are small proteins produced by the phage during its lytic cycle to excise the prophage from the host genome (Lewis and Hatful, 2001). It is not completely understood how RDFs interact with these attachment site-bound integrase dimers to activate *attL-attR* recombination. However, it has been shown that some RDFs (e.g. ϕ C31 integrase and its RDF gp3) do not bind directly to the attachment site DNA, suggesting that they rather interact with the integrase to alter its conformation and activate excision (Khaleel *et al.*, 2011). Studies on ϕ C31 integrase and Bxb1 integrase and their RDFs (gp3 and gp47, respectively) have revealed that these RDFs inhibit formation of *attP-attB* synaptic complexes as well as activating synapsis of *attL-attR* recombination (Ghosh *et al.*, 2006; Khaleel *et al.*, 2011).

As described in Figure 1.6, it is thought that the coiled-coil motifs of integrase dimers, bound to *attL* and *attR* sites following the integration reaction, interact to prevent re-formation of the synaptic complex and therefore inhibiting the reverse *attL-attR* reaction. Rutherford and Van Duyne (2014) suggest that RDFs interact with the coiled-coil motif of the C-terminal domain to end inhibition of synapsis and activating recombination between *attL* and *attR* sites. Their model of this is shown in Figure 1.7. Evidence for this was published by Rowley *et al* (2008), who showed that ϕ C31 integrase mutants with substitutions in the coiled-coil motif were able to catalyse *attL-attR* recombination without the presence of the ϕ C31 RDF, gp3, suggesting that structural changes in this domain activate *attL-attR* recombination.

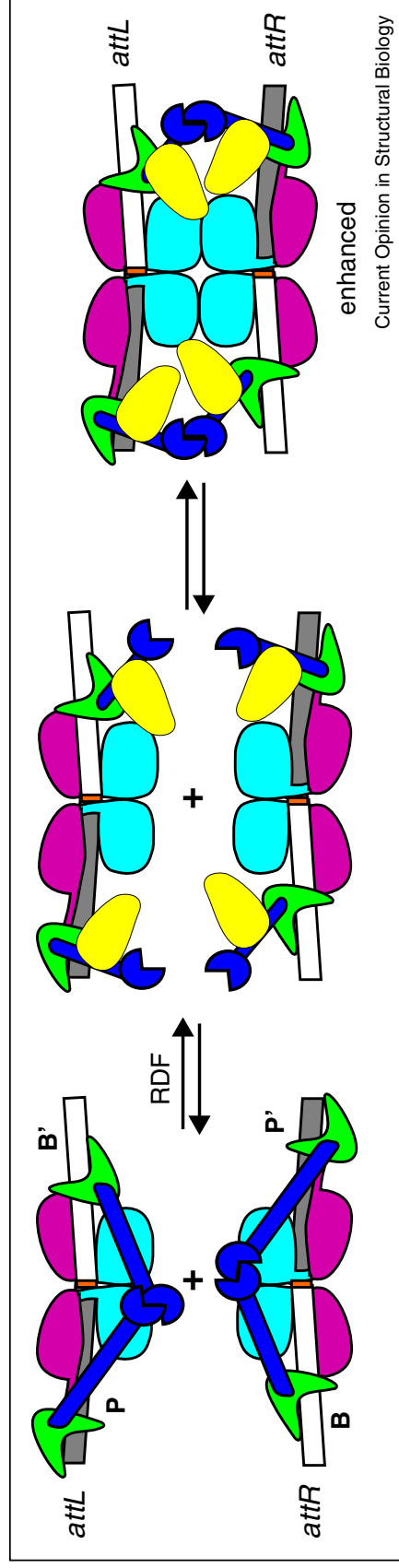


Figure 1.7: Suggested model of integrase-RDF interaction. RDF units (yellow) interact with the coiled-coil motif (blue) of the zinc ribbon domain to alter its conformation in a way that stops inhibition of *attL-attR* recombination and promotes formation of the synaptic complex in which the reverse integration (excision) reaction occurs between *attL* and *attR*. Image: Rutherford and Van Duynne.

1.3.1.4 Attachment site sequence requirements

As the integrase of mechanism of recombination relies on the exchange of two double-stranded DNA half-sites with 2 bp overhangs, it is a requirement of serine integrase recombination that the 2 bp overhangs of recombining *attP* and *attB* sites are complementary. If the central dinucleotide pairings of these sites are not the same, then ligation cannot occur after strand exchange, and the recombination mechanism cannot be completed. Work done by Ghosh *et al* (2003) on Bxb1 integrase revealed the importance of this central 2 bp pairing in the control of directional rearrangements. They suggested that orientation of recombination can be altered by changing this dinucleotide sequence as it does not affect binding to the *att* half-sites, but has a consequence on religation (Ghosh *et al*, 2003). This rule is integral to this project as it allows the use of multiple attachment site sequences, where only sites with matching central dinucleotide pairings are able to recombine. For example, an *attP* site with a TT central overhang can recombine with an *attB* site with a TT overhang, but is unable to recombine with an *attB* site containing a TC overhang. These combinations are shown further in Figure 1.8.

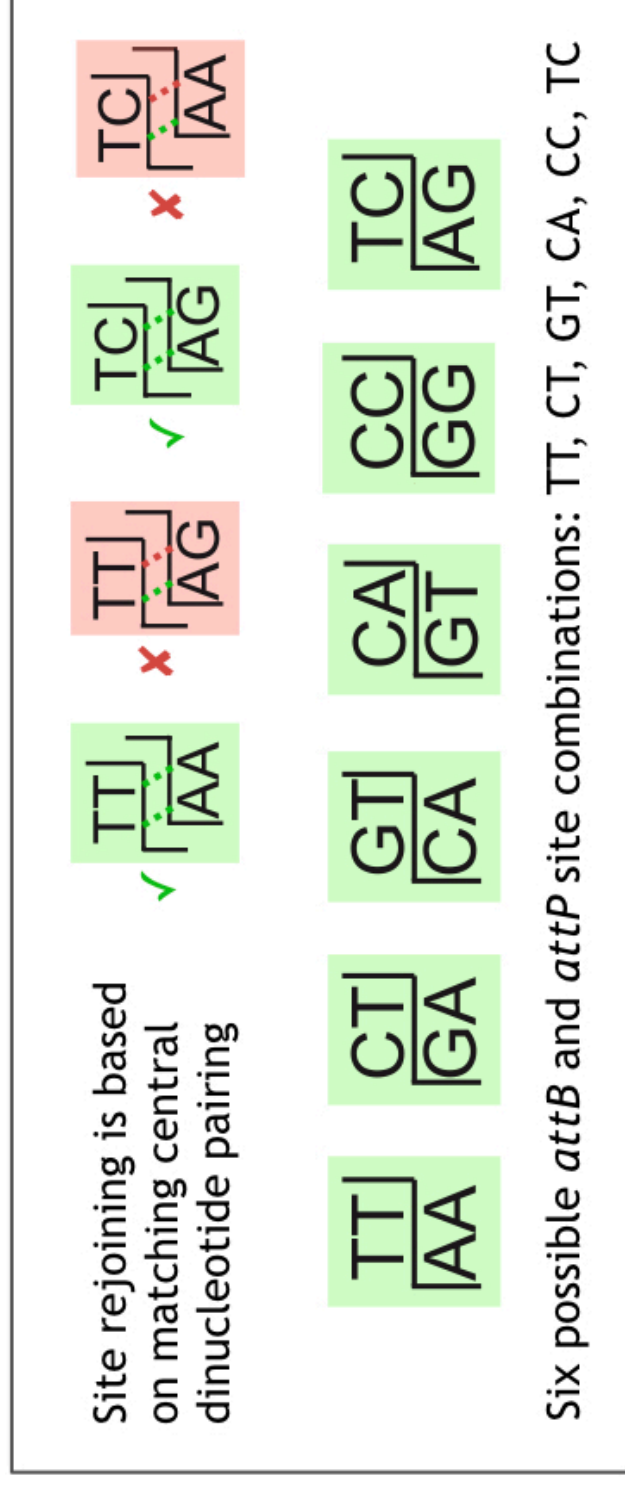


Figure 1.8: The central dinucleotide sequences of recombining serine integrase *att* sites must correspond in order for overlap re-ligation to occur in the synaptic complex during recombination. Shown here are the six possible variants on the central dinucleotide sequence that allow for orthogonal pairing of *att* sites (Colloms *et al*, 2013).

1.3.1.5 Potential applications in synthetic biology

Serine integrases have several characteristics which make them ideally suited to synthetic biology applications. Firstly, they have been shown to function in a highly unidirectional manner, proceeding with the reverse excision reaction between *attL* and *attR* only in the presence of the integrase's RDF. This means that integration reactions are highly efficient, and sets them apart from other enzymes such as the tyrosine recombinases, which are not as strongly unidirectional and therefore less efficient.

Secondly, the minimum attachment site sequences of both *attP* and *attB* required for efficient recombination by serine integrases are much shorter than those of other recombinases proposed for use in synthetic biology (see Table 1.1 for serine integrase minimum attachment site lengths). λ integrase for example requires a much larger *attP* site of 240 bp, despite having a small 25-bp *attB* target site (Landy, 1989). These shorter attachment site pairs are appealing for synthetic biology applications as they shorten the length of required constructs and leave shorter 'scar' sites following recombination. Lastly, unlike tyrosine recombinases, serine integrases do not require auxiliary host factors to complete efficient recombination, further simplifying their use for synthetic biology in a range of organisms. Table 1.1 shows six commonly studied serine integrases that are considered possible candidates for use in synthetic biology, alongside their RDFs, and the minimum attachment site lengths required for efficient recombination.

Table 1.1: Six serine integrases considered candidates for use in synthetic biology, their RDFs, and the minimum *attB* and *attP* sites required for recombination. Here, minimum means the smallest site allowing recombination to occur and not necessarily the most effective attachment site sequence.

Integrase	RDF	<i>attB</i> (bp)	<i>attP</i> (bp)
φC31	gp3 *	34	39
TG1	gp25	39	43
Bxb1	gp47	38	48
φBT1	gp3 *	36	48
φRv1	Rv1584c / Xis	40	52
TP901-1	ORF7	31	50

The potential of serine integrases is expanded further due to the many different combinations of *att* sites can be utilised. Of the several serine integrases that are being used, each recognises its own *att* sites with high specificity. Then, for each integrase there are 6 possible exclusive *attP* and *attB* pairings, based on changes to the central 2 bp sequence (TT, CT, GT, CA, CC, TC). The site-specific nature of these enzymes and the array of possibilities of combination that attachment site modification can enable controlled switching, excision or cassette exchange of DNA on a host's genome.

In addition to the already growing list of serine integrases that are available for use in synthetic biology, chimeric integrases have also been shown to be functional in *E. coli* and human cells. In 2014, Farruggio and Calos created and successfully demonstrated ϕ BT1- ϕ C31 and ϕ C31-TG1 hybrids, which recombined hybrid *att* sites of both integrases, as well as native *att* sites in the case of ϕ BT1- ϕ C31 chimeras (Farruggio & Calos, 2014). While this was the first study to demonstrate successful activity of serine integrase hybrids, it also opens up the possibility of generating many more chimeras which could also be used in a possible multi-integrase synthetic biology toolkit, targeting newly created hybrid sites to further expand the list of *att* site combinations available.

1.3.2. Current uses

1.3.2.1. Serine integrase recombinational assembly (SIRA)

Serine integrases are already being applied in synthetic biology for recombinational plasmid and pathway assembly. The serine integrase recombinational assembly (SIRA) method makes use of the specificity of serine integrases for recombining only sites with matching central dinucleotide pairings to assemble multiple DNA fragments in a one-pot reaction (Colloms *et al.*, 2013). According to the six possible central dinucleotide pairings possible (TT, CT, GT, CA, CC, TC), up to six gene fragments flanked by different *attB* and *attP* sites can be assembled into a circular plasmid in a single reaction with one integrase. Development of this method to incorporate multiple integrases in a one-pot assembly is also further expanding the number of parts it is possible to assemble at once.

1.3.2.2. Serine integrase-mediated genomic integrations

Of focus in this study is the use of serine integrases for rapidly and efficiently integrating heterologous DNA into a host organism's genome. Due to the unidirectional nature of serine integrases and their high specificity for small attachment sites, they are ideally suited for use in genomic integrations. Most genome editing research using serine integrases in heterologous organisms has focused on mammalian cells. Several serine integrases have been demonstrated to effectively insert plasmid DNA into either natural *att*-like pseudo-sites found on mammalian genomes or pre-inserted *att* sites, including ϕ C31 integrase (Thyagarajan *et al.*, 2001), Bxb1 integrase (Russel *et al.*, 2006), ϕ BT1 integrase (Xu *et al.*, 2008) and R4 integrase (Olivares *et al.*, 2001).

However, despite extensive demonstration of the utility of serine integrases for genomic integrations in mammalian cells, there have been comparatively few applications of these enzymes for genome engineering in bacteria. In 2009, Morita *et al* first demonstrated intramolecular recombination in *E. coli* cells using the serine integrase TG1 to recombine TG1 *attP* and *attB* sites across a plasmid substrate *in vivo* (Morita *et al.*, 2009). Following this, they developed a TG1 integrase-based system for the genomic integration of a plasmid in *E. coli*, by first inserting either an *attP* or *attB* site into the host genome and demonstrating integration of a non-replicating plasmid containing the corresponding *att* site under expression of TG1 integrase. They did this first with a 2 kbp size plasmid (Hirano *et al.*, 2011) and later with a much larger plasmid of 10 kbp in size (Muroi *et al.*, 2013) with a high efficiency of $\sim 10^4$ transformants per μg of DNA, showing that increasing sizes of DNA can be integrated with little loss of efficiency.

1.4. A serine integrase-based genomic landing pad system

1.4.1. Landing pad design

This work focused on the design of a novel system in which two serine integrase attachment sites were introduced into the *E. coli* genome as an insert featuring an antibiotic resistance marker gene; this construct is referred to here as a ‘landing pad’. This differs from previous studies in which only a single attachment site on the genome was used to integrate whole plasmid DNA. While the latter is useful for some modes of study, in synthetic biology it is desirable to insert a heterologous gene (or multi-gene construct) into the genome at a known location and in the desired orientation without leaving behind unnecessary markers or other plasmid vector genes.

The landing pad system in this project was designed to incorporate a marker gene flanked by two integrase *attB* target sites with alternate central dinucleotides. A ‘donor’ cassette featuring the gene to be incorporated flanked by the corresponding *attP* sites can then be introduced concomitant with the expression of the integrase enzyme to facilitate cassette exchange. A diagram illustrating this landing pad system is shown in Figure 1.9. The decision to differ the central dinucleotides of each of these sites is based on the requirement of *att* sites to have complementary 2 base overhangs for re-ligation following cleavage and strand exchange in the synaptic complex. For example, a TT central dinucleotide *attB* will only successfully recombine with a similarly TT central dinucleotide *attB*, and the same for TC, CT, GT, CA and CC (see Figure 1.8). There are therefore six useful variants of each *attB-attP* combination for every integrase. Therefore, the result of using two non-identical *attB* sequences as part of the landing pad is that cassette exchange occurs efficiently only in the orientation specified.

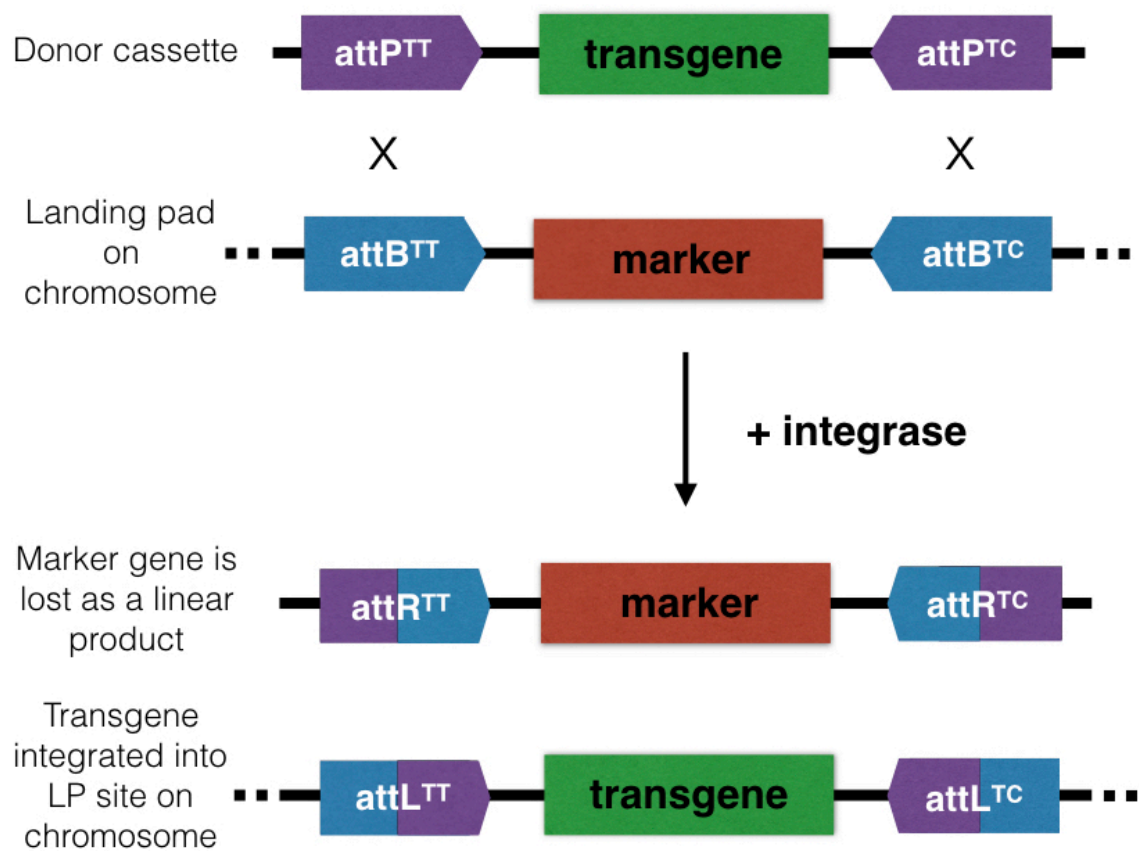


Figure 1.9: The landing pad (LP) system consists of a marker gene flanked by two *attB* sites with non-identical central dinucleotides, *attB^{TT}* and *attB^{TC}*, pre-established on the chromosome via homologous recombination. Under expression of the integrase enzyme, a donor cassette featuring the transgene to be introduced onto the genome flanked by *attP* sites corresponding to the LP *attB* sites (*attP^{TT}* and *attP^{TC}*) is introduced. Recombination between the matching *att* sites results in cassette exchange at the LP site, leaving the transgene on the genome, flanked by recombination product sites *attL^{TT}* and *attL^{TC}*. In this double recombination event the marker gene is removed from the genome, alongside the corresponding recombination product sites *attR^{TT}* and *attR^{TC}*. See Figure 1.2 for clarification of product site formation.

1.4.2. Potential for highly efficient markerless genomic integrations

Given the high efficiency of cassette exchange that could be possible using the landing pad system, there is potential for this mechanism to incorporate target genes without the need for an accompanying marker gene, therefore eliminating the need for a selective marker screening step and expediting the strain development process. If efficiencies prove to be high enough, it may be possible to transform a target gene donor cassette without a marker gene into the landing pad cells while expressing the integrase enzyme and assume that the colonies obtained from this transformation will have performed successful cassette exchange, placing the target gene in the known LP location. For confirmation of the integration, the known genomic location could simply then be sequenced to check for the desired outcome. In this system, the primary limitation would be the initial integration of the LP *attB* sites, which must be done using less efficient homologous recombination methods and requiring markers. However, once LP sites have been established, these strains could be used confidently for integration of markerless cassettes.

1.4.3. Towards a multi-integrase toolkit for whole pathway integration and strain optimisation

Due to the increasing list of integrases available for use in synthetic biology (see Table 1.1), there is an opportunity to create an expanded toolkit of multiple integrase landing pads. A single strain of *E. coli* or other industrially relevant microorganism could be created with a landing pad for each integrase, providing the opportunity to integrate multiple genes of an entire biochemical pathway at separate genetic loci. On top of this, three versions of the landing pad for each integrase could theoretically be created, featuring alternative combinations of *att* sites with two of the six possible central dinucleotide pairings (Figure 1.8). At each site the reverse cassette exchange reaction can also be performed, in which a donor cassette featuring two *attR* sites corresponding to the product *attL* sites of the original cassette exchange is introduced in the presence of the integrase and its RDF. This further increases the options and flexibility of this system for strain development and pathway optimisation. This secondary LP cassette exchange is described in Figure 1.10. However, for this tool to be developed successfully, several questions must be addressed in order to optimise efficiency of cassette exchange at the LP, establish several well characterised LP sites across the genome and ensure that several integrases could operate within the same system without cross-reactivity.

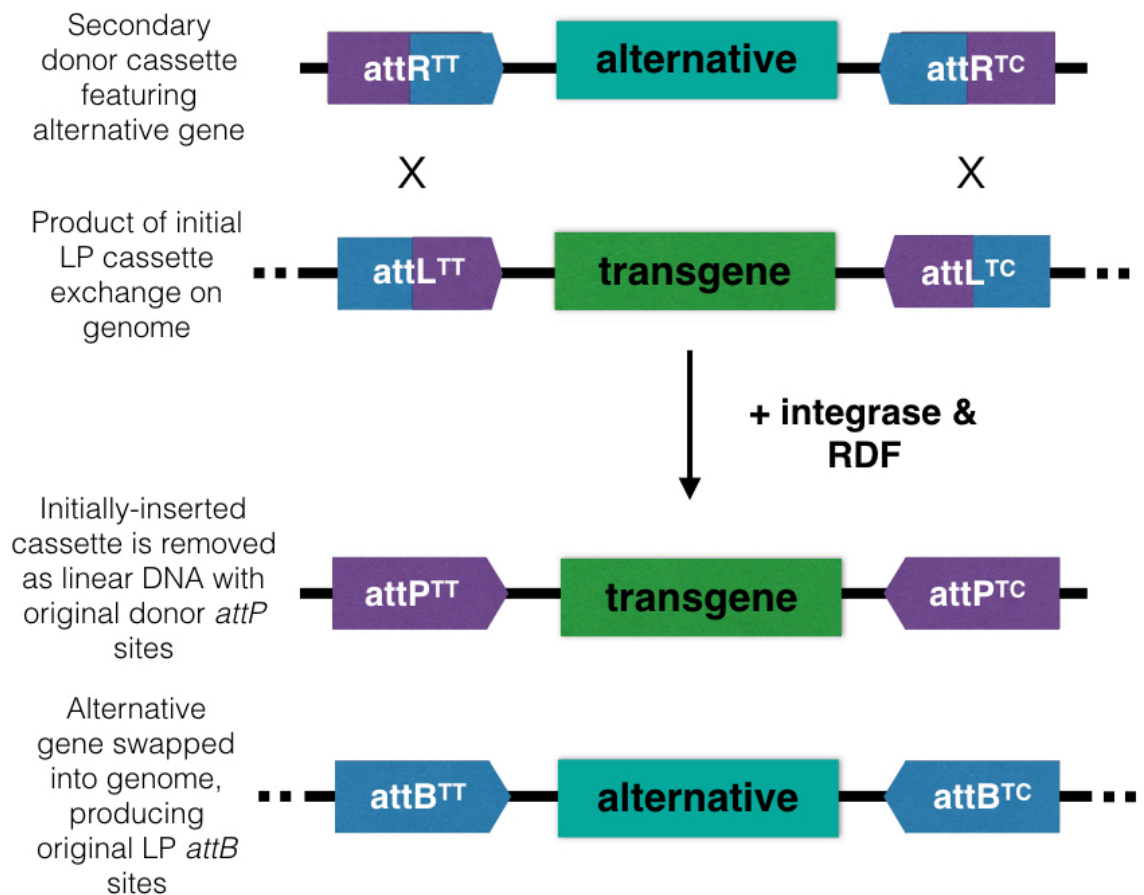


Figure 1.10: Following initial cassette exchange of a heterologous gene into the LP site, a secondary cassette exchange reaction can be performed to replace the initial transgene with an alternative gene. This is done by introducing a new donor cassette featuring the alternative gene flanked by *attR* sites which correspond to the initial LP integration product *attL* sites. Under expression of both the integrase and its RDF, the reverse recombination excision reaction occurs between *attL* and *attR* sites, leaving the new alternative transgene on the genome flanked by *attB* sites. These *attB* sites can then be used again for cassette exchange of an *attP*-featuring donor cassette, continuing the flexibility provided by the LP system.

1.5. Challenges for developing an effective multi-integrase landing pad system

1.5.1. Determining the effect of genome position on landing pad integrations

For a markerless landing pad system to be effective, high cassette exchange efficiencies of close to 100% are needed in order to be confident that a transgene introduced as a donor cassette has been successfully integrated without the need for the subsequent selective screening for which a marker is required. There is already some evidence that position on the *E. coli* genome affects the efficiency of integration of DNA at an introduced *att* site for TG1 integrase (Muroi *et al.*, 2013). Muroi *et al* (2013) pre-inserted *attP* sites at various sites across the *E. coli* genome, and were able to integrate an *attB*-containing plasmid of around 10 kbp in size at an efficiency of close to 10^4 transformants per μg of DNA. They found that integration occurred more efficiently at sites inserted closer to the *E. coli* replication origin site *oriC* and the *E. coli* centromere analogue *migS*, suggesting that this was due to DNA proximal to these locations being exposed more to the cytosol at the poles of the cell during chromosomal segregation. Frequencies of integrations were 10^{3-4} per μg of transformed plasmid for the sites in this region (*dinD* - *yabP*) compared to 10^2 per μg of DNA elsewhere on the genome. Therefore it will be interesting to see whether the efficiency of integration at an LP site is increased further when present in this part of the genome. To investigate this, LPs should be tested at many sites across the *E. coli* genome, in mutants generated by random transposition of the *attB* LP cassette.

Using this set of *E. coli* LP mutants, the variability in expression levels of the transgene could then also be measured at each different site, providing information that could allow an LP mutant to be selected for heterologous gene integration based on the final expression levels required of the product. Genomic location has been shown to have considerable effect on gene expression levels in *E. coli*. Previously, Englaender *et al* (2017) showed that expression levels differed from 25% to 500% of the expression levels of the same gene from a high-copy number plasmid across four different integration sites on the *E. coli* genome. Previous reports support this, with Bryant *et al* (2014) also showing that position on the chromosome can have a 300-fold effect on expression levels of a GFP reporter. A similar measure of expression, using GFP, can be used here to measure expression levels at each randomly-transposed LP site after integration of a GFP-featuring integration cassette. Orientation of a gene on the genome has also been shown to have an effect on expression levels (Block *et al*, 2012), and this will also be taken into account when contrasting expression levels of an integrated GFP gene at each randomly transposed LP site on the *E. coli* genome.

1.5.2. Methods for target gene delivery to the landing pad

The method for delivery of a transgene to be integrated into the genome is a challenge that will be addressed in this project. In initial iterations of phage recombinase-mediated genomic integrations, whole plasmids were integrated at single target sites. Following on from this, plasmid-based cassette exchange methods were then used such as RMCE, in which two non-identical attachment sites flanked a donor cassette on a plasmid, inserting the target gene into a corresponding two-site target cassette on the genome in the desired orientation (Turan *et al.*, 2011).

The most desirable system for delivery of a target gene is via a linear PCR product. This would allow a simple system in which a target gene to be introduced could be amplified by PCR with the two flanking attachment site sequences incorporated via the PCR primers, before transforming these cells with this PCR product. However, transforming cells with linear DNA poses problems due to the activity of endonucleases in the *E. coli* cell such as the RecBCD helicase-nuclease which degrades linear DNA (Dillingham & Kowalczykowski, 2008). Therefore, this project will first devise and implement a method of delivery for the landing pad 'donor cassette' in order to demonstrate the landing pad system, before later going on to focus on possible protocol for delivery via a linear donor cassette in the form of a PCR product.

1.5.3. Orthogonality of serine integrases and their RDFs

In order to develop a multi-integrase landing pad system, there must be minimal cross-reactivity between the different integrases, their RDFs and attachment sites. A multi-LP cell would have sites for several integrases, and therefore for a transgene to be integrated into the correct LP location as desired, there must be a confidence that an integrase will not recombine the sites of another integrase. This could give rise to the transgene being unintentionally integrated into more than one site on the genome. Therefore the fidelity of each integrase to its own attachment sites, the extent to which central dinucleotide pairs direct correct recombination and the orthogonality of integrases and their RDFs will all be investigated, beginning with the latter.

The LP system allows for flexible switching of genes in the genomic LP site, which can be of use when optimising a pathway or testing multiple versions of a possible heterologous gene. For example, following the integration of an initial transgene into an LP, an alternative or mutated gene could then be swapped in directly to replace the initially introduced gene. This could be of advantage for quickly testing different combinations of genes within a pathway during strain expression optimisation. Once the original donor cassette has been integrated at the LP site via recombination of *attP* and *attB* sites, two non-identical *attL* sites (with differing central dinucleotide pairs) are left on the genome flanking the integrated gene. These *attL* sites can be used in another cassette exchange reaction, integrating a donor cassette flanked by corresponding *attR* sites. This then leaves behind two *attB* sites as they existed on the original LP cassette. A diagram of this secondary LP cassette exchange reaction is shown in Figure 1.10.

In a multi-integrase system, the success of this chop-and-change mechanism depends on the fidelity of an integrase with its RDF. If an integrase was expressed in the cell for the ‘forward’ exchange reaction (*attB* x *attP*) when residual compatible RDF was present, cassette exchange efficiency could in theory be reduced as the RDF would activate the reverse reaction and the integrase would therefore no longer have the advantage of unidirectional efficiency. For the development of a successful multi-integrase landing pad system, the extent to which chosen integrases are cross-compatible in the above ways must be determined. Creating a protocol for sequential cassette exchanges at several integrase LPs will also be a major challenge, as it will require the expression and subsequent suppression of each integrase and RDF for secondary cassette switching.

1.6. Project aims

The foremost aim of this project was to demonstrate a landing pad for ϕ C31 integrase in *E. coli*. This involved designing the landing pad, inserting it into a defined location on the genome and measuring the efficiency of cassette exchange at this landing pad site. This work makes up the content of Chapter 3 of this thesis. Specific objectives include:

- Designing a landing pad cassette for ϕ C31 integrase and selecting an appropriate location on the genome into which it will be inserted.
- Synthesising a cassette containing this LP sequence with flanking homology to the target site.
- Inserting the ϕ C31 integrase into the *E. coli* genome via recombineering.
- Confirmation of insertion of the ϕ C31 LP into the genome by selective screening and colony PCR.
- Design and construction of a 'donor cassette', and development of an effective delivery method that allows for detection transformation and cassette exchange.
- Measurement of cassette exchange efficiency at this ϕ C31 LP site. Note: cassette exchange efficiency is defined throughout this thesis as the percentage of cells (colonies) in which cassette exchange reaction has taken place across the LP, as a proportion of the number of cells that have been successfully transformed with the donor cassette DNA (colonies selecting positive for donor cassette transformation).
- Establish and test a Bxb1 integrase LP in the same design as the ϕ C31 integrase LP.

After the ϕ C31 integrase LP has been demonstrated in a single location on the genome, LPs will be established at multiple locations on the *E. coli* genome in order to measure the relative efficiency of different genomic loci, with a view to development of a multiplexed landing pad system. This work makes up the content of Chapter 4 of this thesis. Specific objectives include:

- Inserting the ϕ C31 integrase LP at several locations on the *E. coli* genome by random transposition.
- Confirming and characterising the genomic locations of a section of the ϕ C31 LP mutant library generated by random transposition, and selecting several ϕ C31 LP mutants to represent a diversity of LP locations across the entire genome.
- Measuring the efficiency of cassette exchange at each of these selected ϕ C31 LP strains to determine whether genomic location has an effect on recombination and LP cassette exchange efficiency.
- LP sites can be further characterised by measuring expression levels at each genomic locus, by inserting a fluorescent protein gene into the LP via cassette exchange and measuring fluorescence.

Another aim of the project was to determine the extent to which closely related integrases and their RDFs are capable of interacting. ϕ C31 integrase and TG1 integrase, which share a high percentage of sequence homology, will first be tested to see whether they interact with each integrase's RDFs. This work makes up the content of Chapter 5 of this thesis. *In vitro* recombination assays, in which a plasmid substrate containing two corresponding attachment sites is incubated with the integrase (with or without RDF), with recombination resulting in the production of two smaller plasmids, will be used to investigate the following:

- To what extent TG1 RDF gp25 is capable of activating ϕ C31 *attL-attR* recombination
- To what extent ϕ C31 RDF gp3 is capable of activating TG1 *attL-attR* recombination
- To what extent TG1 RDF gp25 is capable of inhibiting ϕ C31 *attP-attB* recombination
- To what extent ϕ C31 RDF gp3 is capable of inhibiting TG1 *attP-attB* recombination

The final aim of the project was to export the ϕ C31 integrase LP demonstrated in *E. coli* into the simple eukaryotic host *Saccharomyces cerevisiae*. This involves designing host-appropriate selections to replace the one used in *E. coli*, but with the same overall landing pad design. Using a mutant knockout yeast library, ϕ C31 LPs will be established in several locations on the *S. cerevisiae* genome, and the relative cassette efficiencies between each will be measured. This work makes up the content of Chapter 6 of this thesis.

2. Chapter 2: Materials and Methods

Chapter 2 details the methods used in this study. Section 2.1 to Section 2.5 list the microbial strains, growth media, antibiotics, PCR primers and plasmids used or generated during this work. Sections 2.6 to 2.22 are divided into individual protocols used and are described to ensure reproducibility and clarity of experimental method.

2.1. Bacterial and yeast strains

All *Escherichia coli* strains were derived from strain K-12 substrain MG1655. Reference names and genotypes of each strain used in this study are shown in Table 2.1. The *Saccharomyces cerevisiae* strains used in this study are derived from strain S288C, and are shown in Table 2.2.

Table 2.1 : *Escherichia coli* strains used in this study. All stored as glycerol stocks (50% v/v) at - 70°C.

<i>E. coli</i> Strain Name	Genotype	Reference
DS941	AB1157, <i>recF</i> , <i>lacZ</i> ΔM15, <i>lacI</i> ^q	Summers & Sherratt, 1988
π1	DH5α <i>thyA::(erm-pir116)</i> (<i>ermR</i>)	Demarre <i>et al</i> , 2005
LPC311	AB1157, <i>recF</i> , <i>lacZ</i> ΔM15, <i>lacI</i> ^q , <i>pepA::φC31LP(ermR)</i>	This study
RTC31LP	AB1157, <i>recF</i> , <i>lacZ</i> ΔM15, <i>lacI</i> ^q <i>φC31LP(kanR,rpsL)</i>	This study

Table 2.2 : *Saccharomyces cerevisiae* strain used in this study. Library stored in multi-well plate glycerol stocks (50% v/v) at - 80°C.

<i>S. cerevisiae</i> Strain Name	Genotype	Reference
YKO (<i>KanMX</i>)	Background strain <i>Saccharomyces cerevisiae</i> S288C	Giaever & Nislow, 2014

2.2. Growth media

All growth media used in this study are listed in Table 2.3.

Table 2.3 : Liquid and solid (agar) media used in this study

Name of Media	Composition
Liquid LB	10 g tryptone, 5 g yeast extract, 5 g NaCl; made up to 1 L with distilled H ₂ O and adjusted to pH 7.5 with NaOH
LB Agar	Liquid LB plus 15 g / L agar
MacConkey Agar	8 g MacConkey agar base, 180 mL distilled H ₂ O, 10 mL 20% galactose (plus antibiotics as required)
Synthetic Defined (SD) Yeast Media	2% glucose, 6.7 g/L yeast nitrogen base, amino acids (20 mg/L histidine, 120 mg/L leucine, 60 mg/L lysine, 20 mg/L arginine, 20 mg/L tryptophan, 20 mg/L tyrosine, 40 mg/L threonine, 20 mg/L methionine, 50 mg/L phenylalanine, 20 mg/L uracil, 20 mg/L adenine), [20 g/L agar for solid media]: add distilled H ₂ O to 1 L and sterilised by autoclave
Yeast Extract Peptone Dextrose (YPD)	10 g yeast extract, 20 g bacto peptone, 20 g dextrose, [24 g agar for solid media]: added distilled H ₂ O to 1 L and sterilise by autoclave

2.3. Antibiotics

Antibiotics were used throughout this study to select for the following antibiotic-resistance makers on plasmids or cassette inserts: *ApR* (ampicillin-resistance), *KmR* (kanamycin-resistance), *CmR* (chloramphenicol-resistance) and *EmR* (erythromycin-resistance). The antibiotics used with their working and stock concentrations are shown in Table 2.4.

Table 2.4 : Antibiotics used in this study with working concentrations used for antibiotic marker selections in growth media and stock concentrations.

Antibiotic	Working Concentration	Stock Concentration
Ampicillin	100 µg / mL	100 mg / mL in H ₂ O
Kanamycin	50 µg / mL	50 mg / mL in H ₂ O
Chloramphenicol	25 µg / mL	25 mg / mL in 100% EtOH
Erythromycin	200 µg / mL	50 mg / mL in 100% EtOH

2.4. PCR Primers

The majority of primers used in this work were synthesised as oligonucleotides by Eurofins Genomics. For PCR and sequencing primers under 50 bp the following Eurofins synthesis specifications were used: Unmodified DNA Oligo, 0.05 µmol synthesis scale, with High Purity Salt Free (HPSF) purification. Larger primers (over 50 bp) were synthesised by Integrated Gene Technologies (IDT) as Ultramer DNA Oligos, at a scale of 4 nmol, with High Performance Liquid Chromatography (HPLC) purification. All primers used are shown in Table 2.5 in order of use throughout this work.

Table 2.5 : Primers used in this study

Name	Size (bp)	Sequence (5' to 3')	Application
C31LP pepA For	123	ATTCTATCTGTAGCCACCGCCGTTGTC TTTAAGATTCAGGAACGTAGTGCGTG CGGGTGCCAGGGCGTGCCCTTGGGC TCCCCGGGCGCGTACTCCCTTAGAAG CAAACCTTAAGAGTGTGTTG	Amplification of PepAC31LP recombineering DNA fragment
C31LP pepA Rev	129	AATAGTTATCTATTATTTAACGGGAGG AAATAAGGAGTACGCGCCCGGGGAGC CCGAGGGCACGCCCTGGCACCCGCA CAAACCGCGCTGGGTTTAACGGCGAA GAGTAATTGCGTCAGGCAAGGCTGT	Amplification of PepAC31LP recombineering DNA fragment

Name	Size (bp)	Sequence (5' to 3')	Application
Upstr. pepA For	20	ACACGAAGTCATCGCAACAG	Confirmation PCRs at <i>pepA</i> /LPC31 site (recombineering and cassette exchanges)
Dwnstr. pepA Rev	20	CTCATCGCCTGTGAAGATGA	LP confirmation PCRs (see above)
pepA For	18	GGAGAAACAGCGGAGTGC	LP confirmation PCRs (see above)
pepA Rev	19	AGCGTCTCTTTTGCCGTCT	LP confirmation PCRs (see above)
EmR For	15	GGAGGGATTCGTCAT	LP confirmation PCRs (see above)
EmR Rev	15	TATGGCGGGTAAGTT	LP confirmation PCRs (see above)
CmR Rev	20	GGGCGAAGAAGTTGTCCATA	LP confirmation PCRs (see above)
KmR For	20	TTGGGTGGAGAGGCTATTCG	LP confirmation PCRs (see above)
KmR Rev	20	AATATCACGGGTAGCCAACG	LP confirmation PCRs (see above)
C31LP Donor For	90	CGGCGGAAGCTTGTAGTGCCCCAACT GGGGTAACCTTTGAGTTCTCTCAGTT GGGGGCGTAGGCTAGCAAGAGGTTC CAACTTTCACCAT	Amplification of <i>KmR</i> / <i>CmR</i> ϕ C31LP donor cassette for pRM1/2/3

Name	Size (bp)	Sequence (5' to 3')	Application
C31LP Donor Rev	97	GCAGTTATTGGTGCCTAGAAAGCGGC CGCCGGCCGCTACGCCCCCACTGAG AGAACTCGAAGGTTACCCAGTTGGG GCACTACGGTACCGCGGCC	Amplification of <i>KmR/CmR</i> ϕ C31LP donor cassette for pRM1/2/3
C31LP EmR/ NheI Don For	37	CCCCCGCTAGCGAATTCCTTAGAAGC AAACTTAAGAG	Amplification of an <i>EmR</i> ϕ C31LP donor cassette for pRM4
C31LP EmR/ NotI Don Rev	39	AAAAAGCGGCCGCCTCGAG TTATTCCTCCCGTTAAATA	Amplification of an <i>EmR</i> ϕ C31LP donor cassette for pRM4
KAN3	21	GACCGCTTCCTCGTGCTTTAC	Arbitrary PCR (Methods 2.19)
KAN4	23	TCTATCGCCTTCTTGACGAGTTC	Arbitrary PCR (Methods 2.19)
ARB1	35	GGCCACGCGTCGACTAGTCANNNNN NNNNNGATAT	Arb. PCRs (N = random base)
ARB2	20	GGCCACGCGTCGACTAGTCA	Arbitrary PCR (Methods 2.19)
RTL3 For	20	GTGCTGTCGTTTCATACCGAC	Confirmation of RT LP 3 location
RTL5 For	20	ACATAGCCCTCATCTACCGC	Confirmation of RT LP 5 location
RTL7 For	20	TGGCTGTCTTTCGAACCTCT	Confirmation of RT LP 7 location
RTL8 For	20	GTGCTGGATGGGCTGTATTG	Confirmation of RT LP 8 location
RTL9 For	20	CTTAACGCCTCTTCATCCGC	Confirmation of RT LP 9 location

Name	Size (bp)	Sequence (5' to 3')	Application
RTL10 For	20	CCTTCCAGTGACCATGTTGC	Confirmation of RT LP 10 location
M13 Uni (-21)	18	TGTAACGACGGCCAGT	Sequencing plasmids
M13 Rev (-29)	18	CAGGAAACAGCTATGACC	Sequencing plasmids
Yeast LP 1F	20	GCGCGTACTCCACCTTCTAG	Confirmation of KanMX::URA3LP
Yeast LP 2F	25	GAAACATGAAATTGCCAGTATTCT	Confirmation of KanMX::URA3LP
Yeast LP 1R	20	GGCCGCAGGTGGAGTACG	Confirmation of KanMX::URA3LP
Yeast LP 2R	25	TATTGTTGGAAGAGGACTATTTGCA	Confirmation of KanMX::URA3LP

2.5. Plasmids

Plasmids created during this study were made through a combination of restriction digest and ligations from existing plasmids and PCR products. Further details on these methods can be found later in this chapter. All plasmids produced and used during this study are detailed in Table 2.6. The six plasmids constructed as part of this work are detailed further in Figures 2.1 to 2.6.

Table 2.6: Plasmids used in this study

Plasmid	Antibiotic Marker	Description	Source
pRM1	<i>ApR</i>	φC31LP donor temperature-sensitive (<i>repA101ts</i>) plasmid; <i>CmR</i> cassette	This study (Figure 2.1)
pRM2	<i>CmR</i>	φC31LP donor temperature-sensitive (<i>repA101ts</i>) plasmid; <i>KmR</i> cassette	This study (Figure 2.2)
pRM3	<i>ApR</i>	φC31LP donor temperature-sensitive (<i>repA101ts</i>) plasmid; <i>KmR</i> cassette	This study (Figure 2.3)
pRM4	<i>ApR</i>	φC31LP donor temperature-sensitive (<i>repA101ts</i>) plasmid; <i>EmR</i> cassette	This study (Figure 2.4)
pRM5	<i>ApR</i>	φC31 yeast LP delivery vector (homologous recombination); <i>URA3</i> cassette marker; <i>E. coli</i> ColE1 vector	This study (Figure 2.5)
pFEM32	<i>ApR</i>	φC31 integrase expression plasmid (constitutive)	Femi Olorunniji
pFEM35	<i>ApR</i>	Bxb1 integrase expression plasmid (constitutive)	Femi Olorunniji
pKOBEG-A	<i>ApR</i>	λ Red (<i>gam</i> , <i>bet</i> , <i>exo</i>) expression for recombineering; <i>pBAD</i> ; <i>RepA101ts</i>	Chaverroche <i>et al.</i> , 2000
pKOBEG-C	<i>CmR</i>	λ Red (<i>gam</i> , <i>bet</i> , <i>exo</i>) expression for recombineering; <i>pBAD</i> ; <i>RepA101ts</i>	Chaverroche <i>et al.</i> , 2000

Plasmid	Antibiotic Marker	Description	Source
pSW23	<i>CmR</i>	Plasmid used as template for PCR amplification of ϕ C31LP donor cassette marker (see pRM1)	Demarre <i>et al</i> , 2005
pSW29	<i>KmR</i>	Plasmid used as template for PCR amplification of ϕ C31LP donor cassette marker (see pRM2 & RM3)	Demarre <i>et al</i> , 2005
pGD001	<i>KmR</i>	<i>Galk</i> (MacConkey) assay substrate; ϕ C31 integrase <i>attP</i> & <i>attB</i> sites	Femi Olorunniji
pZJ7	<i>CmR</i>	ϕ C31 integrase expression under pBAD (arabinose-induced); p15A ori	Jia Zhao
pZJ73	<i>CmR</i>	ISY100 plasmid for random transposition of <i>KmR</i> ϕ C31 LP into the <i>E. coli</i> genome	Jia Zhao
pFM16	<i>ApR</i>	<i>in vitro</i> recombination substrate; ϕ C31 integrase <i>attP</i> & <i>attB</i> sites	Femi Olorunniji
pFM52	<i>ApR</i>	<i>Galk</i> (MacConkey) assay substrate; ϕ C31 integrase <i>attL</i> & <i>attR</i> sites	Femi Olorunniji
pFM138	<i>ApR</i>	<i>Galk</i> (MacConkey) assay substrate; TG1 integrase <i>attP</i> & <i>attB</i> sites	Femi Olorunniji
pFM139	<i>ApR</i>	<i>Galk</i> (MacConkey) assay substrate; TG1 integrase <i>attL</i> & <i>attR</i> sites	Femi Olorunniji
pMTL23	<i>ApR</i>	<i>E.coli</i> cloning vector used for creation of yeast LP plasmid pRM5 (see Figure 2.5); ColE1 origin	Chambers <i>et al</i> , 1988

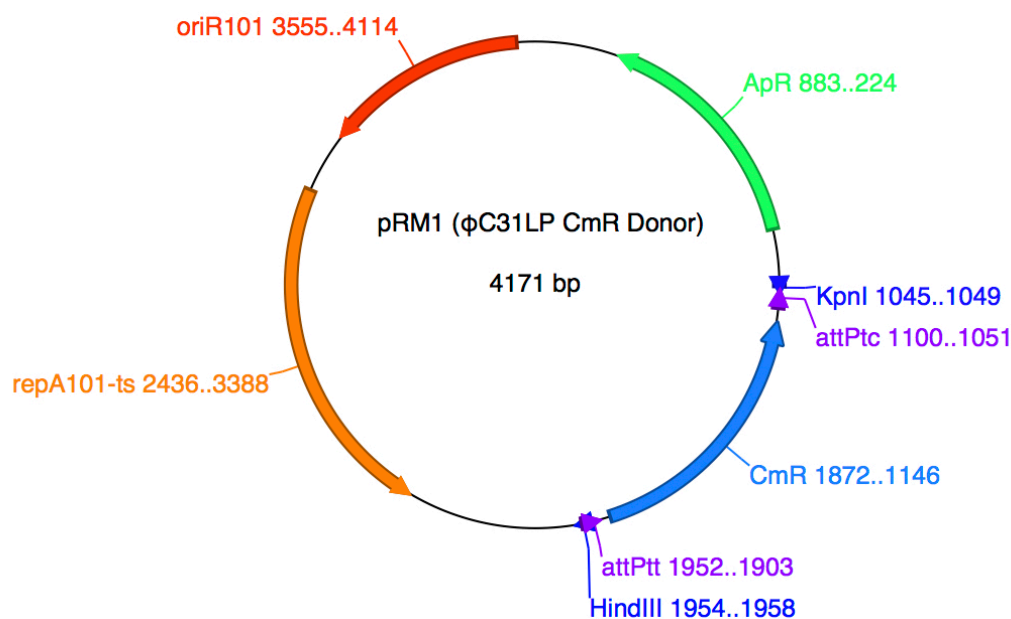


Figure 2.1: pRM1, constructed in this study. This φC31 donor cassette (*CmR* flanked by *attP^{TT}* and *attP^{TC}*, PCR amplified from pSW23 using primers C31LP_Don_For & _Rev) was cloned into pKOBEG-A between restriction sites *KpnI* and *HindIII*. pRM1 features *repA101-ts* for temperature sensitivity (see Section 3.3.1).

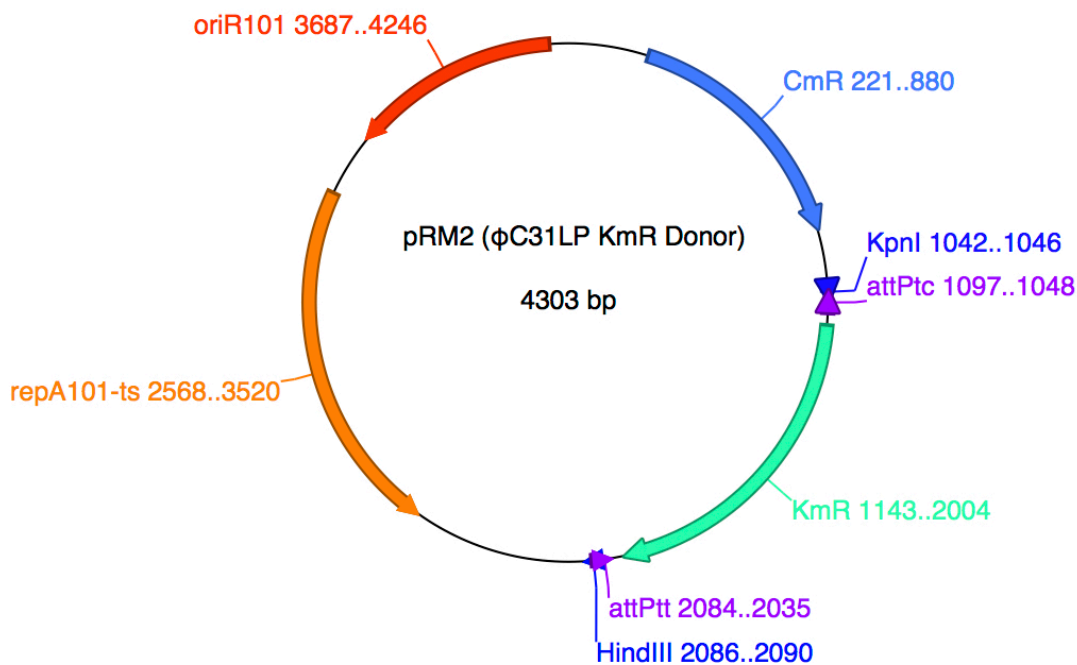


Figure 2.2: pRM2, constructed in this study. This φC31 donor cassette (*KmR* flanked by *attP^{TT}* and *attP^{TC}*, PCR amplified from pSW29 using primers C31LP_Don_For & _Rev) was cloned into pKOBEG-C between restriction sites *KpnI* and *HindIII*. pRM2 features *repA101-ts* for temperature sensitivity (see Section 3.3.1).

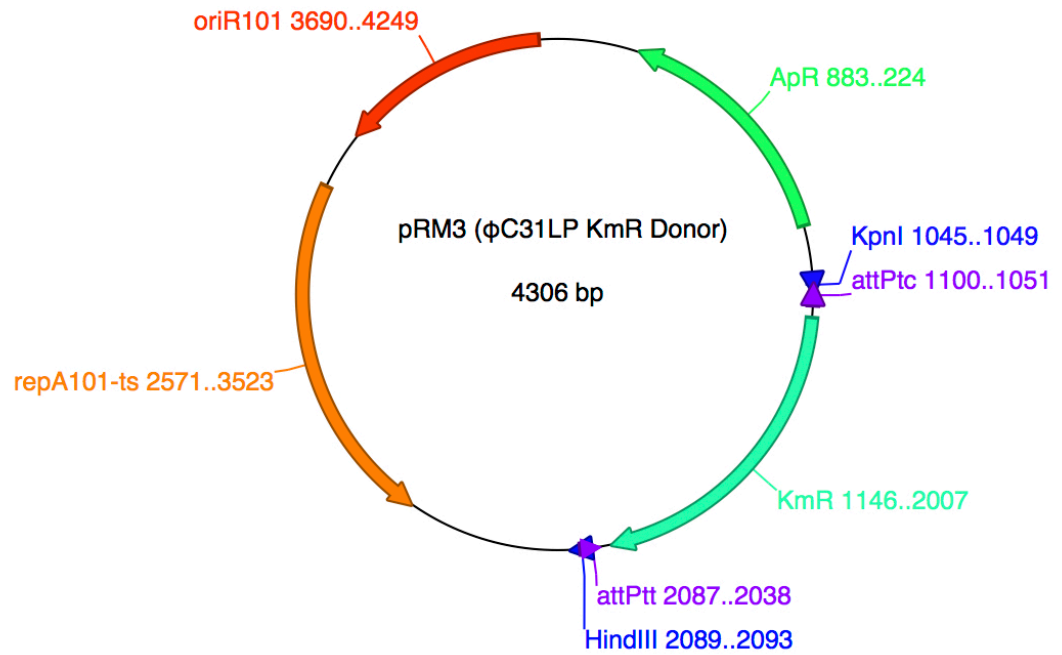


Figure 2.3: pRM3, constructed in this study. This φC31 donor cassette (*KmR* flanked by *attP^{TT}* and *attP^{TC}*, PCR amplified from pSW29 using primers C31LP_Don_For & _Rev) was cloned into pKOBEG-A between restriction sites *KpnI* and *HindIII*. pRM3 features *repA101-ts* for temperature sensitivity (see Section 3.3.1).

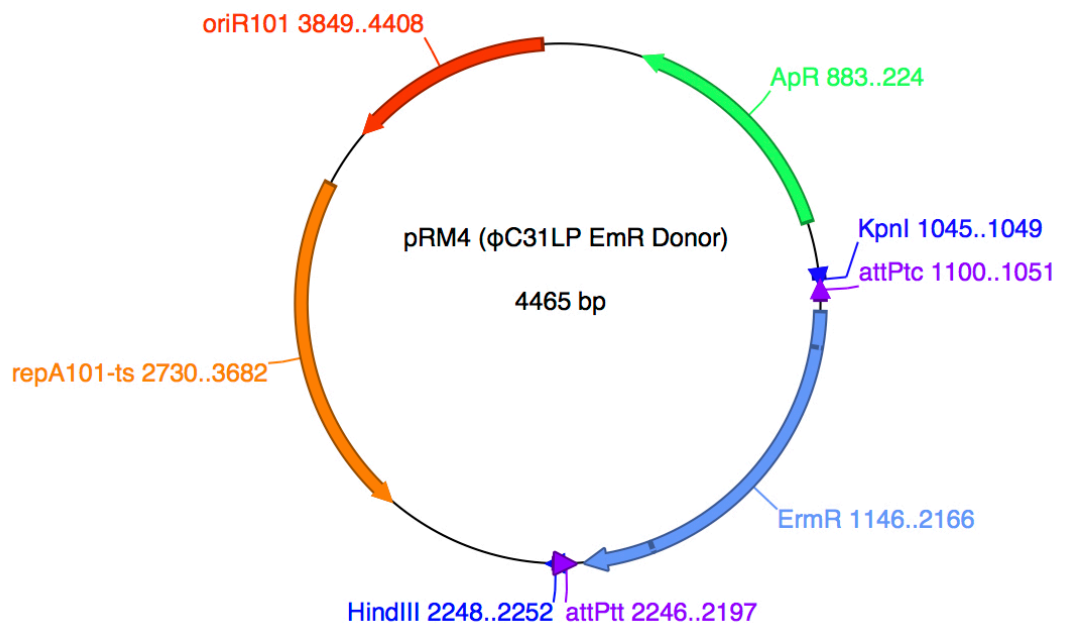


Figure 2.4: pRM4, constructed in this study. This φC31 donor cassette (*CmR* flanked by *attP^{TT}* and *attP^{TC}*, PCR amplified from pSW23 using primers C31LP_Don_For & _Rev) was cloned into pKOBEG-A between restriction sites *KpnI* and *HindIII*. pRM4 features origin *repA101-ts* for temperature sensitivity (see Section 3.3.1).

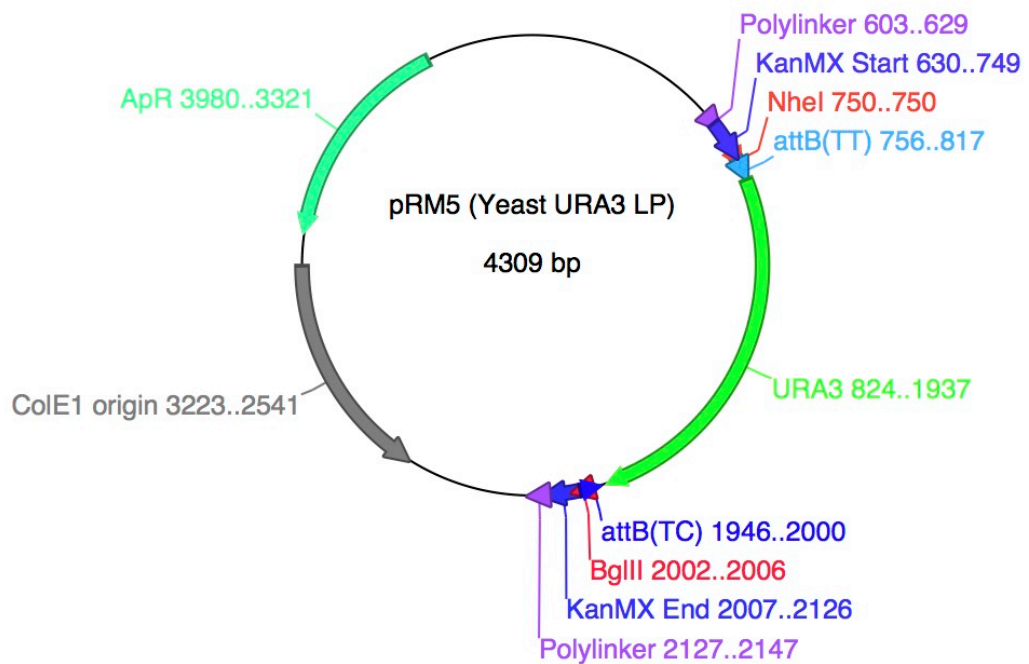


Figure 2.5: pRM5, constructed in this study. This plasmid features the yeast ϕ C31 LP consisting of a *URA3* marker gene flanked by attachment sites *attBTT* and *attBTC*, surrounded by 150 base pairs of homology to the start and end of the *KanMX* gene for replacement by homologous recombination. The plasmid has a ColE1 origin for cloning and propagation in *E. coli* prior to transformation into yeast.

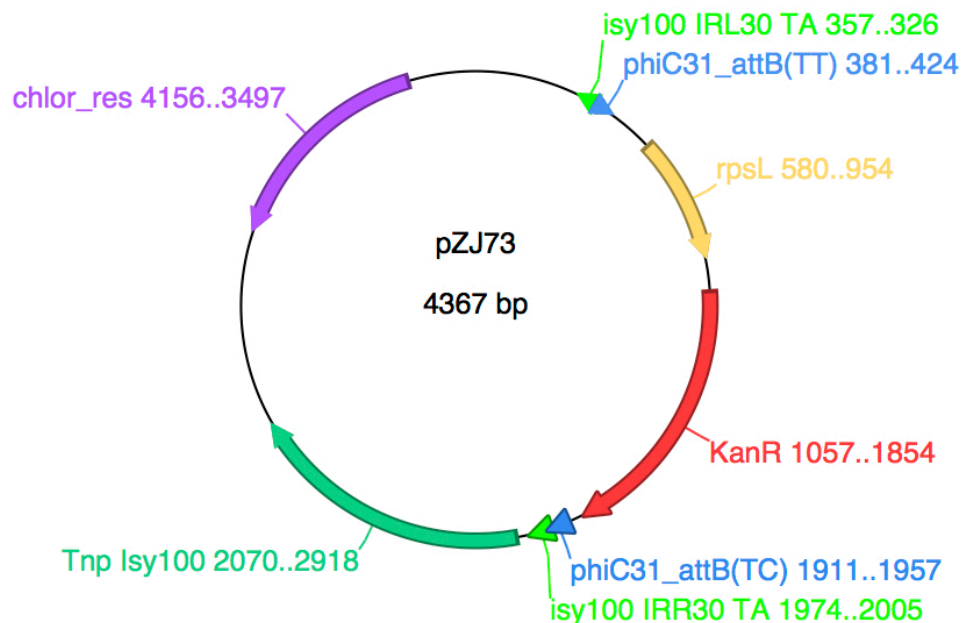


Figure 2.6: pZJ73 was the plasmid used to generate the *E. coli* DS941 ϕ C31 LP mutant library used in Chapter 5 via ISY100 random transposition. The ϕ C31 LP, consisting of a kanamycin resistance gene and counter-selectable *rpsL* marker flanked by *attBTT* and *attBTC* sites, placed between the two ISY100 terminal inverted repeat sequences IRL and IRR, targets and inserts into the *E. coli* genome at any TA sequence (Urasaki *et al*, 2002).

2.6. Restriction endonuclease digests

Most digests were done using NEB endonucleases according to the manufacturer-recommended protocol. Digests were set up in microcentrifuge tubes containing the sample DNA in double distilled H₂O, supplemented with the endonuclease manufacturers recommended buffer, with 2-10 units of enzyme per microgram of DNA. Samples were incubated for at least 1 hour in a water bath pre-heated to 37°C.

2.7. Ligation of plasmid DNA

To ligate restriction digested fragments (from plasmid or PCR digests) during the cloning of new plasmids, the following standard protocol was used for ligation reactions with T4 ligase (NEB). In a microcentrifuge tube, linear digested DNA fragments were mixed with the NEB manufacturers 10x concentrated T4 buffer (volume for 1x final concentration) and made up to volume of 20 µL with double distilled H₂O. The ligation mixture was incubated at room temperature (around 16°C) overnight. The following morning, 1-2 µL of this ligation mixture was used directly for transformation into electrocompetent *E. coli* cells, while the rest was saved for DNA purification by ethanol precipitation.

2.8. Ethanol precipitation

The following ethanol precipitation protocol was used to purify and isolate DNA following ligations for plasmid cloning. To an overnight ligation reaction mixture, 0.22 volumes 5M ammonium acetate was added, followed by 2.5 volumes 100% (v/v) ethanol, and vortexed vigorously before freezing overnight at -20°C. The DNA was then pelleted by centrifugation at 13.2k rpm in a 4°C centrifuge for 60 minutes. The supernatant was very carefully removed, and the pellet washed in 80% (v/v) ethanol before centrifuging again for 30

minutes. The supernatant was then removed and the DNA pellet dried in a vacuum evaporator for 5 to 10 minutes. 20 μ L of double distilled H₂O was added to the pellet and vortexed vigorously before spinning down at 10k rpm for 2 minutes and cooling on ice for 5 minutes. This solution could then be stored at -20°C or used immediately for transformation into electrocompetent *E. coli* cells.

2.9. Standard transformations of *E. coli* cells

2.9.1. Chemical transformation

For standard chemical transformation of plasmid DNA into *E. coli* cells, the following protocol was used. 0.4 mL of overnight growth culture was added to 20 mL of sterile L-broth (LB) in a sterile 100 mL volume conical flask. This was incubated at 37°C for 90 minutes before being transferred to a centrifuge tube. Cells were centrifuged at 4°C at 7000 rpm for 4 minutes and the supernatant carefully discarded. 10 mL of ice-cold CaCl₂ was added to the tube and the pellet carefully resuspended on ice. The CaCl₂ cell resuspension was left on ice for one hour before repeating the above centrifugation step and discarding the supernatant. The remaining cell pellet was resuspended in 1 mL CaCl₂, and the chemically competent cells left on ice until needed for transformation.

For each DNA transformation, 100 μ L of the chemically competent cells was placed in a pre-chilled microcentrifuge tube, to which 1-2 μ L of DNA was added, with flicking to mix. The cells and DNA were incubated on ice for 20 minutes before being heat-shocked in a 37°C water bath for 5 minutes and replaced on ice for another 5 minutes to cool. 200 μ L LB was added to the transformation mixture and incubated at 37°C (30°C for temperature-sensitive (TS) plasmids) for 2 hours with shaking at 225 rpm. After this recovery incubation, 100-200 μ L was spread on L-agar plates containing the appropriate antibiotic for selection of the plasmid backbone and the plates incubated overnight at 37°C (30°C for TS) to select for successfully transformed colonies.

2.9.2. Electrotransformation

Throughout this work, a high-efficiency electrotransformation protocol was used based on a transformation method recommended for λ Red recombineering (Chaveroche *et al.*, 2000). Each electrotransformation was done as follows (with changes where noted). A colony of *E. coli* cells was inoculated into 5 mL of LB (with appropriate antibiotic) and grown overnight at 37°C (30°C for TS) with shaking at 225 rpm. In the morning, 0.5 mL of this overnight culture was added to two sterile 100 mL conical flasks containing 25 mL LB (plus antibiotic). This was incubated at 37°C (30°C for TS), shaking at 225 rpm, until an optical density OD₆₀₀ of around 0.4 was reached (around 3 hours of growth). The culture was then chilled in an ice-water slurry for 10 minutes before being transferred to pre-chilled centrifuge tubes and centrifuged at 4°C for 7 minutes at 6500 x g. The supernatant was removed and the cell pellet carefully resuspended in 1 mL of ice-cold 10% glycerol (90% double distilled water, filter sterilised) on ice. Another 25 mL ice-cold 10% glycerol was added and the above centrifuge step was repeated. Cells were again resuspended in 1 mL ice-cold 10% glycerol plus another 25 mL and the centrifuge step repeated a third time. The cell pellet was again resuspended in 10% ice-cold glycerol and transferred to a pre-chilled microcentrifuge tube. This was centrifuged for 30 seconds at maximum speed in a microcentrifuge pre-chilled to 4°C and the supernatant was carefully discarded. Remaining cells were then resuspended in 200 μ L ice-cold 10% glycerol.

For the electrotransformation step, 40 μ L of electrocompetent cells were transferred to a pre-chilled microcentrifuge tube. 1-2 μ L of salt-free DNA was added and mixed by flicking. This mixture was then transferred to a pre-chilled electrotransformation cuvette on ice. Cells were transformed in a MicroPulser electroporator using the EC2 setting (see product specifications for more details). After shocking, 1 mL of LB was immediately added and the broth-cell mix transferred to a sterile microcentrifuge tube. Tubes were incubated at 37°C (30°C for TS) with shaking at 225 rpm for at least 2 hours before spreading 100-200 μ L of recovered cells onto L-agar containing appropriate antibiotics. Plates were then incubated at 37°C (or 30°C for TS) overnight to select for successfully transformed colonies.

2.10. Agarose gel electrophoresis

Agarose gels were used to separate out DNA according to size, from restriction digests or recombination assays. Agarose gels were made by adding agarose to TAE buffer (40 mM Tris/acetate, 1 mM EDTA, pH 8.5) for a final concentration of 1% agarose and heating in a microwave to completely dissolve. The heated liquid agarose was then left to cool before pouring into a gel electrophoresis dock containing a well-forming comb and being left to set. The comb was then removed and the gel submerged in TAE buffer before adding DNA samples (vortex-mixed with 6.25x concentrated loading dye to a final concentration of 1x). The gel electrophoresis was run at a voltage of 80-120 V for 60-120 minutes (or until loading dye had reached an appropriate place on the gel) before stopping and staining the gel.

DNA in the gel was visualised by ultraviolet (UV) light using an ethidium bromide stain. After electrophoresis, gels were added to TAE buffer bath with added ethidium bromide at a 0.6 µg/mL concentration. The gels were then shaken gently in the ethidium bromide TAE solution for 30-60 minutes, before being rinsed and de-stained in TAE for a further 30-60 minutes to remove non DNA-bound ethidium bromide from the gel. DNA was then visualised using UV. For DNA fragments that were to be excised and purified for subsequent *in vitro* or *in vivo* work, agarose gels were placed on a 365 nm wavelength UV transilluminator so as to avoid too much UV damage to the DNA. For photography of the DNA on the agarose gels, a 254 UV transilluminator was used.

2.11. Extraction of DNA from agarose gel

For all DNA extractions from agarose gels, DNA fragments were visualised and excised using a scalpel on a 365 nm wavelength UV transilluminator. DNA was then purified from the excised gel using a QIAgen Gel Extraction Kit according to the manufacturers protocol.

2.12. DNA sequencing

All sequencing was done by MWG Eurofins Genomics. DNA samples were prepared according to MWG Eurofins' recommendations and dried in a vacuum evaporator before posting to Eurofins. If PCR primers were readily available in the lab, primer oligos were mixed with the plasmid samples in the concentration recommended by Eurofins prior to drying. Where possible, both 3' and 5' strands of DNA were sequenced using forward and reverse primers, to doubly check intended ligations or integrations are as intended.

2.13. PCR reactions

2.13.1. NEB Phusion Polymerase method

For PCRs with longer primers (such as the amplification of genes for recombineering featuring long additions of sequence homology), a modified NEB Phusion Polymerase protocol was used. Reactions were set up as follows: in a PCR tube mix 10 µL 5x NEB HF Polymerase Buffer, 5 µL 2mM (each) dNTP mix, 1 µL 50 mM MgCl₂, 1.5 µL DMSO (3% final), 5 µL forward primer (5 µM), 5 µL reverse primer (5 µM), 1 µL template DNA (for amplifying directly from a strain's genome, 1 colony was transferred to 100 µL double distilled water and vortexed vigorously, before adding 1 µL of this to the PCR mix) and 22 µL double distilled water. The above mixture was vortexed before adding 1 µL NEB Phusion polymerase and vortexing again briefly. Reactions were then run through the following heat cycle: 1) 98°C for 60 seconds, 2) 98°C for 20

seconds, 3) 55°C for 30 seconds, 4) 72°C for 90 seconds, 5) repeat steps 2-4 29 more times, 6) 72°C for 10 minutes. Reactions were then cooled at 4°C before either purifying by agarose gel electrophoresis and gel extraction, or using a QIAgen PCR Purification Kit (to avoid degradation by UV radiation).

2.13.2. Taq Polymerase method

For simpler PCRs with shorter primers, an NEB 2x Taq Master Mix protocol was used, set up according to NEB recommendations. For amplifications from genomic DNA (as in cassette exchange confirmation PCRs), 1 colony was transferred to 100 µL double distilled water and vortexed vigorously, before adding 1 µL of this to the PCR mix. The following heat cycle was then used: 1. 95°C for 30 seconds, 2. 95°C for 20 seconds, 3. 55°C for 30 seconds, 4. 68°C for 90 seconds, 5. repeat steps 2-4 24 more times, 6. 68°C for 10 minutes. Reactions were then cooled at 4°C and run on an agarose gel electrophoresis.

2.14. Purification of PCR product

For most experiments, PCR products were purified by first running on an agarose gel, excising the DNA band from the gel on and extracting the DNA from the agarose using a QIAgen Gel Extraction Kit. However, in some cases it was necessary to avoid exposing the DNA to UV radiation (for example in order to improve efficiency of homologous recombination in recombineering), a QIAgen PCR Purification Kit was used. A small amount of the DNA yielded from this purification was then run on an agarose gel electrophoresis to confirm the size of the PCR product.

2.15. λ Red recombineering

The recombineering protocol was primarily based on the high-electrotransformation method described in Methods 2.9.2 (Chaveroche *et al.*, 2000). *E. coli* DS941 was first chemically transformed with the λ Red expression plasmid pKOBEG-A (see Methods 2.5), with all consequent steps done at 30°C to maintain the temperature-sensitive pKOBEG vector.

DS941/pKOBEG-A were grown for further electrotransformation as described in Method 2.9.2 with the addition of arabinose-induction of λ Red elements under control of the pBAD promoter. Cultures were grown at 37°C until they reached an OD₆₀₀ of ~ 0.2, before adding 10% (w/v) L-arabinose (in filter sterilised double distilled water kept at 4°C) for a final concentration of 0.2% w/v. After induction cells were returned to a 37°C shaking incubator until a final OD₆₀₀ of ~0.4 was reached.

100 μ L of high-efficiency electrocompetent *E. coli* DS941/pKOBEG-A cells were then transformed with 4 μ L of the purified recombineering cassette PCR product. For the recombineering of LPC31 (replacing *pepA* with a ϕ C31 integrase *attB* LP with erythromycin-resistance marker *EmR*) this PCR product was amplified from *E. coli* strain π 1 (*EmR*-positive) using primers C31LP_*pepA*_For and C31LP_*pepA*_Rev (Methods 2.4). It is important to note that during the development of this cassette exchange protocol, it was discovered that exposing PCR product to UV radiation (at the time of extraction from electrophoresis agarose gel) resulted in no integration of the PCR product by homologous recombination. Therefore in the final iteration of this protocol the PCR product DNA was purified directly from the PCR mixture using a QIAgen PCR Purification Kit.

2.16. Clearing cells of temperature-sensitive plasmids

Following transformations with temperature-sensitive plasmids (pKOBEG-A for recombineering or pRM plasmids for LP cassette exchanges), cells were cleared of the plasmids in the following method. Marker-positive transformants (selected according to the antibiotic marker being integrated into the genome) were chosen from transformation plates and transferred by toothpick to a 100-grid LB agar plate containing no antibiotic (100 transformants), which was then incubated at 42 °C overnight. The following day, each transformant patch was again transferred by toothpick to another 100-grid LB agar plate containing no antibiotic and again incubated at 42 °C overnight. Following this incubation, each temperature-treated transformant patch was transferred by toothpick to 100-grid agar plates containing different antibiotics to test for presence of the temperature-sensitive plasmid. For example, presence of pKOBEG-A would be tested for on LB agar plates containing ampicillin as this is the backbone marker of this plasmid. Additional selective plates would be selected to test for presence of the marker being integrated into the genome, and in the case of LP cassette exchange, for the marker gene being replaced on the LP.

2.17. Landing pad cassette exchange transformations

Landing pad cassette exchange experiments were based on the high-efficiency electrotransformation protocol described in Method 2.9.2 (Chaverocche *et al.*, 2000). Initial cassette exchange transformations, in which a constitutive ϕ C31 integrase expression plasmid was used, were done as described in Method 2.9.2. However, later cassette exchange protocols were done using an arabinose-inducible ϕ C31 integrase expression plasmid, and followed the alterations described in Method 2.15 (for λ Red recombineering). Recovered cells after transformation were spread onto LB agar plates containing the antibiotic marker on the LP donor cassette, before curing the cells of temperature-sensitive donor plasmids as described in Method 2.16.

2.18. ISY100 random transposition of LP in *E. coli*

The ISY100-generated library of *E. coli* DS941 ϕ C31 LP (*KnR*) mutants was generated in work done previously by this lab (E. Conte, manuscript in preparation). In brief, the ISY100 transposon protocol was done by introducing a plasmid which expresses ISY100 and features the LP cassette flanked by inverted repeat sites IRL30 and IRR30 (see Figure 2.6) into DS941 cells via a standard electrotransformation method. Successful integrations were selected on agar plates containing kanamycin which were then scraped and the resulting cell suspension used to create a glycerol stock library.

2.19. Screening of random LP mutants by arbitrary PCR

The arbitrary PCR method used in this study to confirm locations of randomly transposed ϕ C31 LPs in the *E. coli* genome was adapted from a protocol described by Das *et al* (2005). The confirmations consist of a two-step PCR protocol, described further in Chapter 4. All primers used are described in Methods section 2.4, Table 2.5.

The first PCR, which amplifies from an arbitrary sequence according to randomly generated sequences of primer ARB1 to a sequence within the kanamycin resistance gene of the ϕ C31 LP, primer KAN3, was set up in the following Phusion polymerase reaction: 27.5 μ L double distilled H₂O, 10 μ L 5x Phusion HF buffer, 2 μ L 50 mM dNTPs, 2.5 μ L 5 μ M KAN3, 2.5 μ L 50 μ M ARB1, 5 μ L cell supernatant (1 LP mutant colony suspended in 50 μ L double distilled H₂O and boiled for 15 mins) and 0.5 μ L Phusion polymerase. This reaction mixture was put into the following PCR machine heat cycle: 1) 95°C for 5 minutes, 2) 94°C for 30 seconds, 3) 30°C for 30 seconds, 4) 72°C for 90 seconds, 5) repeat steps 2-4 6 times, 6) 94°C for 30 seconds, 7) 55°C for 30 seconds, 8) 72°C for 2 minutes, 9) repeat steps 6-8 30 times, 10) 72°C for 10 minutes, 11) cooled at 4°C.

The second PCR step, which amplifies from a sequence within the ARB1 primer (primer ARB2) to a sequence within the LP kanamycin resistance gene but slightly downstream of primer KAN3 (primer KAN4), was set up in the following way: 31 μL double distilled H_2O , 10 μL 5x Phusion HF buffer, 2 μL 50 mM dNTPs, 2.5 μL 5 μM KAN4, 2.5 μL 5 μM ARB2, 1.5 μL of the first PCR reaction and 0.5 μL Phusion polymerase. This reaction mixture was put into the following PCR machine heat cycle: 1) 95°C for 5 minutes, 2) 94°C for 30 seconds, 3) 52°C for 30 seconds, 4) 72°C for 90 seconds, 5) repeat steps 2-4 30 times, 6) 72°C for 10 minutes, 7) cooled at 4°C.

Following the second PCR, 5 volumes PB (from a QIAGEN Plasmid DNA Miniprep Kit) was added to the mixture, and the PCR product was cleaned by following the QIAGEN Gel Extraction kit protocol. Final PCR product DNA was eluted in 30 μL double distilled H_2O and sent for sequencing with the primer KAN4.

2.20. *in vitro* integrase reactions

In vitro recombination assays were set up in microcentrifuge tubes as follows: 21 μL IRB3 buffer, 8 μL plasmid substrate, 5 μL integrase enzyme at 8 μM were added to a single tube. For integrase-RDF reactions, a 2.5 μL sample of each component at 8 μM concentration were mixed together (for final concentrations of 4 μM integrase and 4 μM RDF) and incubated at room temperature for 15 minutes prior to addition of buffer and substrate. The sample was then vortexed and incubated in a 30°C water bath for 30 minutes before stopping the reaction by heating in a 80°C heat block for 5 mins followed by cooling on ice for 5 mins. Added to this was 25.5 μL B103 buffer and 49.5 μL *Nrul*. The sample was then incubate at 37°C for 1 hour. Following this incubation, 15 μL of an SDS / protease K / EDTA mix was added and incubated for another 15 mins at 37°C. Samples were then run in an agarose gel electrophoresis and visualised by UV.

2.21. *in vivo galK* recombination assay

In vivo recombination activity was tested using the *galK* MacConkey assay. *E. coli* cells containing an integrase expression plasmid (either continuous or arabinose-induced; see plasmids in Methods Section 2.5) were chemically transformed with the *galK attP/attB* substrate plasmid pGD001 (Methods Section 2.5), with transformants selected on MacConkey agar plates (Methods Section 2.2) containing kanamycin (pGD001 backbone) and an appropriate antibiotic to select for the expression plasmid. If recombination has taken place across the pGD001 substrate, colonies appear yellow/white due to loss of the *galK* gene, but appear pink/red if sufficient recombination has not occurred (this is explained in more detail in Chapter 3).

After photographing plates, cells were ‘scraped’ from the agar by adding 1 mL of LB and pulling an ethanol-sterilised glass spreading rod across the plate to dislodge cells. Plasmid DNA was then extracted from these cells using a QIAGEN Plasmid DNA Miniprep Kit, and the resulting DNA sample was run in an agarose gel.

2.22. Yeast lithium acetate transformations

Yeast KanMX knockouts were first streaked from glycerol stocks onto YPD and incubated for 2 days at 30°C. From single colonies, overnight cultures of 10 mL YPD liquid media were set up at a temperature of 30°C with shaking at 200 rpm. After incubation for 15 hours, the OD₆₀₀ of each was recorded at 0.4 - 0.5, equivalent to 5 x 10⁶ cells per mL. From these cultures, 2 mL was inoculated by sterile pipetting into conical flasks containing 50 mL YPD media. The 50 mL cultures were then incubated at 30°C, shaking 200 rpm, for 3-4 hours until an OD₆₀₀ of 0.1 - 0.2 was achieved, roughly equivalent to 2 x 10⁶ cells per mL. The following lithium acetate transformation method was based on the high efficiency yeast transformation protocol published by Pan *et al.* (2007). Each 50 mL, 4-hour culture was harvested by centrifugation at 5,000 rpm for 2 minutes at room temperature, then washed once in 10 mL of sterile water. Cells were then washed again with 10 mL of 0.1 M lithium acetate, prepared fresh in buffer TE (pH 8.0). Pellet was gently resuspended in residual lithium acetate to a total volume of 300 µL, enough for 3 reactions. For each transformation reaction, 100 µL of this cell suspension was transferred to a fresh 1.5 mL microcentrifuge tube, kept on ice.

A transformation mixture was prepared fresh for each reaction as follows: 480 µL 50% polyethylene glycol (PEG-3350, Sigma) in buffer TE (pH 8.0), 72 µL 1M lithium acetate, 40 µL denatured salmon sperm carrier DNA (ssDNA) at 10 mg/mL (ssDNA placed in a heat block at 100°C from frozen for 5 minutes, then immediately cooled on ice prior to use), 20 µL DNA (~1 µg per reaction) and buffer TE (pH 8.0) to volume.

620 µL of the transformation mixture was added to each 100 µL 0.1M lithium acetate cell suspension and mixed thoroughly by pipetting only. The reaction was then incubated at 30°C for 30 minutes in a water bath before adding 72 µL dimethyl sulfoxide (DMSO) and mixed thoroughly by pipette. Cells were then heat shocked at 42°C for 15 minutes in a water bath, then harvested by centrifugation at 3,600 rpm for 30 seconds. Cells were resuspended in 400 µL 5mM CaCl₂ and incubated at room temperature for 10 minutes. Following this,

cells were spun down at 3,600 rpm for 30 seconds and resuspended in 1 mL YPD liquid media. Microcentrifuge tubes containing the resuspended cells were then placed horizontally in a 30°C incubator shaking at 200 rpm for 2.5 hours, then harvested by centrifugation at 3,000 rpm for 2 minutes. Pelleted cells were finally resuspended in 200 µL YPD liquid media and plated onto SD-uracil agar plates for URA selection using a sterilised glass rod. Transformation plates were incubated at 30°C for 2 - 3 days.

2.23. DNA quantification from agarose gels (digital)

To roughly quantify the relative concentrations of DNA bands in agarose gels from recombination assays, the software ImageJ (developed by the National Institutes of Health) was used to analyse gel photographs. Firstly, gel images were black-white inverted. In ImageJ, rectangular areas were selected at the four expected band locations in each column on the gel. Each column (i.e. each recombination assay reaction) was analysed individually. From these four areas, ImageJ measures the relative pixel concentrations in each band, first plotting the data as peaks. Each peak was isolated from the next using the line drawing tool. ImageJ then calculates the relative intensity/area of each peak as a percentage. The percentage of substrate DNA that had been recombined or unrecombined could then be calculated by adding together the two recombined product bands and the two unrecombined bands. See Chapter 5 for more details.

3. Chapter 3: Demonstrating an integrase ‘landing pad’ (LP) in *E. coli*

3.1. Introduction

The integrase landing pad (LP) system on which this project focuses has been designed to facilitate integration of a target gene into *E. coli* or other microbial genomes with maximum efficiency and accuracy. The LP cassette consists of a marker gene flanked by two *attB* sites in a head-to-head orientation. In this set-up, the two *attB* sequences differ by their central dinucleotide pairings, making use of the requirement for matching attachment sites described previously in Chapter 1. This feature ensures that a ‘donor’ cassette, consisting of a target gene flanked by the corresponding *attP* sites in a head-to-head orientation, is integrated in the desired orientation and at high efficiency due to site-directed cassette exchange.

An early study which demonstrated how attachment sites on the genome could be used for heterologous gene integration was done by Bierman *et al* (1992), who site-specifically integrated heterologous plasmid DNA containing a single *attP* site into the ϕ C31 native host *Streptomyces*. Single attachment sites have also been inserted, or natural pseudo-sites discovered, in eukaryotic cells which allow integrations of plasmid DNA through a single *attB* x *attP* site recombination reaction (Thyagarajan *et al.*, 2001). However, the LP system described here marks an improvement on these previous methods as it allows for integration of a gene at a chosen location, wherever an LP sequence is introduced, and requires only one integration event to place the target gene on the genome, without the need for consequent excision of a plasmid backbone. The efficiency and accuracy of the LP system is also expected to be higher than previous single-site methods as it requires the site-specific recombination of two differing *attB-attP* pairings in the correct orientation, reducing the likelihood of off-target integration events.

The initial target of this project was to demonstrate a landing pad with sites for ϕ C31 integrase on the *E. coli* genome. ϕ C31 integrase has been well characterised by this lab and has been demonstrated to work efficiently in *E. coli* without detriment to cell growth.

This section of the project aims to establish a ϕ C31 integrase LP in a chosen location on the *E. coli* genome via an intermediate recombineering step, and demonstrate ϕ C31 integrase-mediated cassette exchange at this site.

Cassette exchange experiments were designed to measure efficiency of exchange using loss and gain of antibiotic resistance markers. When a marker on the genomic LP site is lost and a donor cassette marker is gained in a successful cassette exchange, this can be tested in the resulting colonies by checking for antibiotic resistance or sensitivity to the associated antibiotics.

A significant challenge was to come up with a way to deliver the donor cassette while still allowing consequent selection of the antibiotic resistance markers. Plasmid delivery is the most stable and effective method of delivery of the donor cassette. However, as plasmids remain in the cells after transformation, presence of the donor marker (and plasmid backbone marker) poses a problem for selection screening. The plasmid used must therefore be removed from the cells before checking for antibiotic resistances of the LP and donor cassettes.

The ideal delivery method for this system is the transformation of a simple PCR product, as discussed previously. While this would avoid the issue of marker incompatibility and unwanted persistence in the cells, linear DNA suffers from problems of instability in *E. coli* cells due to action of native endonucleases. Alongside these challenges, an high efficiency cassette exchange protocol must also be optimised for simultaneous donor delivery and integrase expression and activity.

3.2. Design and construction of a ϕ C31 integrase landing pad

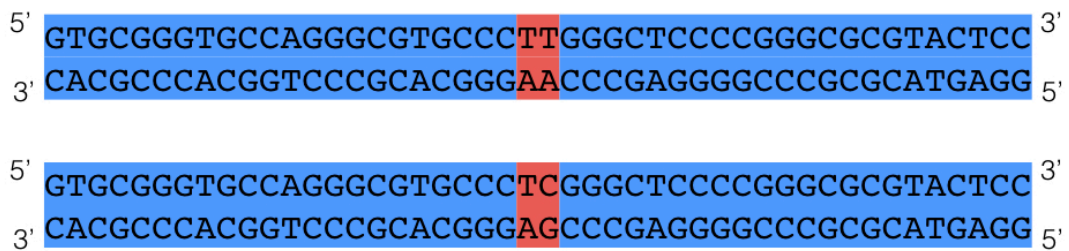
3.2.1. Design of the ϕ C31 LP construct

For the first landing pad design of this project, two *attB* sites for ϕ C31 integrase were arranged in a head-to-head orientation around the antibiotic resistance marker for erythromycin, *EmR*. An overview diagram of this initial ϕ C31 integrase landing pad is shown in Figure 3.1. The first *attB* site features a TT central dinucleotide pair and the second *attB* site features a TC central dinucleotide pair; these *att* sites are shown in Figure 3.2. The sequences, which are smaller than the complete ϕ C31 integrase *att* site sequences found naturally, were chosen to be shorter where possible while still maintaining the highest possible level of recombination efficiencies. The sites selected (a 46 bp *attB* site and 50 bp *attP* site) are based on a previous study (Groth *et al.*, 2000) which showed that reducing the length of a natural *attB* site (60 bp from *Streptomyces lividans*) to 35 bp in length had no effect on recombination efficiency, while reducing the natural ϕ C31 integrase *attP* site (60 bp) to 46-50 bp in length also produced recombination efficiencies equal to that of the full sites. The central dinucleotide pairs of TT (which is unchanged from the natural ϕ C31 integrase *att* sites), and TC were chosen due to observations in this lab that altering the TT to a TC does not have an effect on recombination efficiency, and that these sites have been shown to recombine exclusively with their matching central dinucleotide counterpart site.



Figure 3.1: Initial ϕ C31 integrase landing pad design consisting of an erythromycin resistance marker flanked by attachment sites $attB^{TT}$ and $attB^{TC}$. Sequences of the $attB$ sites in this design are shown in Figure 3.2.

46 bp ϕ C31 $attB$ sites



50 bp ϕ C31 $attP$ sites

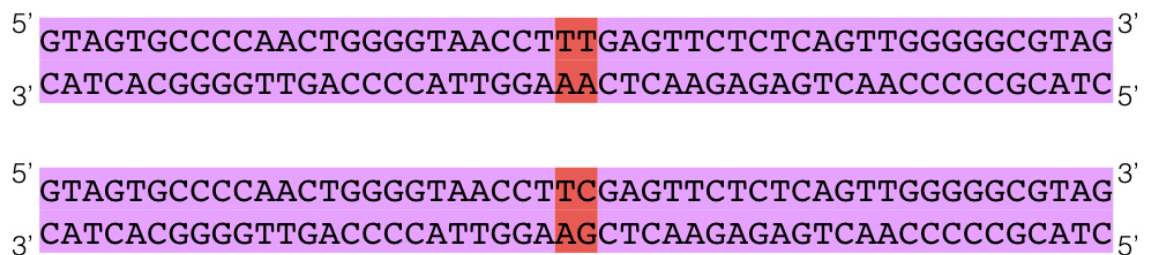


Figure 3.2 : Sequences of ϕ C31 integrase $attB$ and $attP$ sites used in the ϕ C31 landing pad (see Figure 3.1) and donor cassettes (see Figure 3.6), with alternate central dinucleotide pairings TT and TC.

3.2.2. Insertion of the LP cassette into the *E. coli* gene *pepA* by λ Red recombineering

The location for this initial genomic landing pad was selected as the non-essential *E. coli pepA* gene, due to it having been previously mutated without negative impact to the cell (Reijns *et al.*, 2005) and the availability of a *pepA* knockout assay. The *pepA* gene encodes an aminopeptidase, which has a role as an accessory protein in the XerC/XerD recombinase dimer resolution mechanism (Reijns *et al.*, 2005).

The ϕ C31 integrase landing pad was inserted into the *pepA* locus, replacing the entire gene, via the λ Red recombineering method. This is a method of facilitating integration of a PCR product into the genome via homologous recombination. The desired sequence to be integrated is amplified by PCR with flanking regions of homology to the targeted gene. The PCR product is then introduced into *E. coli* cells expressing the λ Red proteins, which block activity of cellular endonucleases and promote homologous recombination. A diagram illustrating how this mechanism acts to replace the *pepA* gene with a homology-added ϕ C31 integrase LP (*EmR*) PCR product is shown in Figure 3.3.

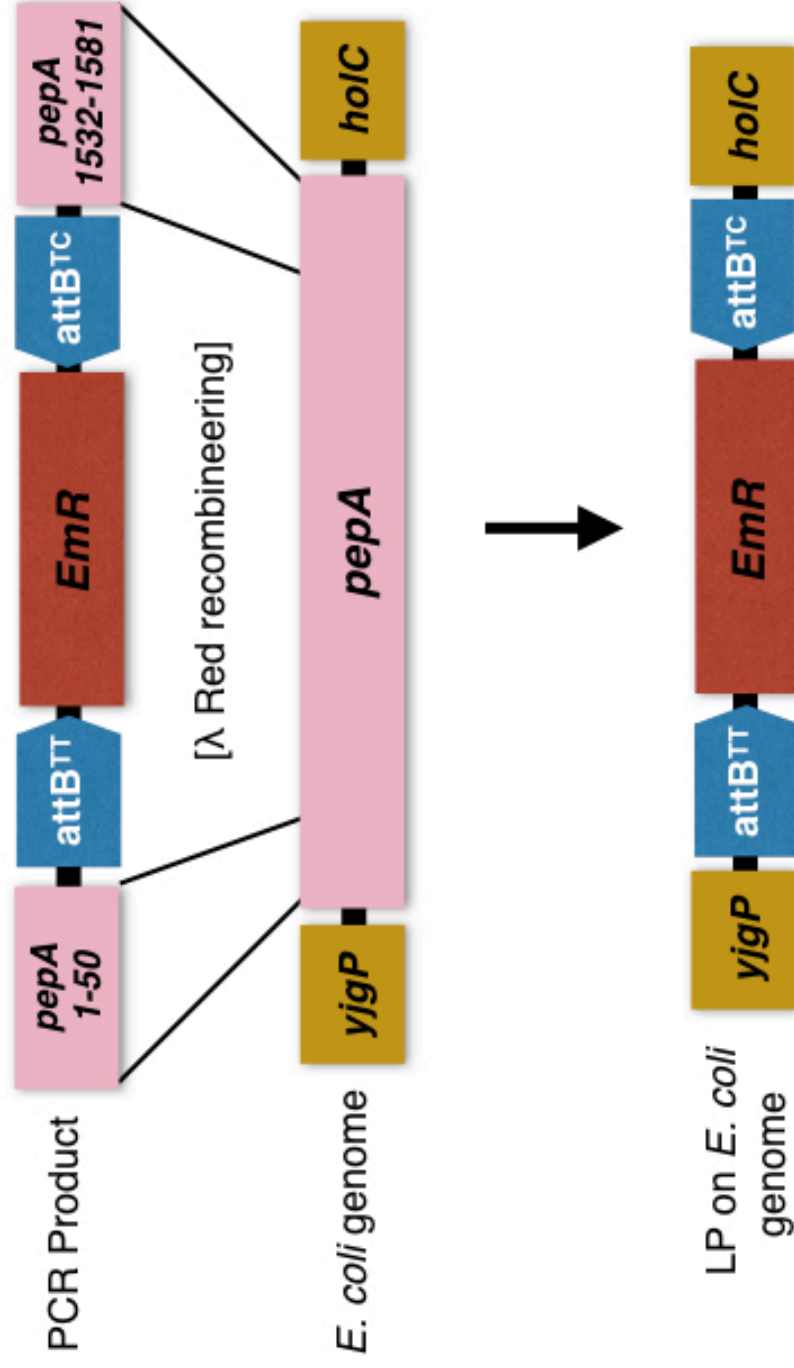


Figure 3.3: Using the λ Red recombineering method, a PCR product consisting of the ϕ C31 integrase LP (*EmR*) flanked by homology sequences to the start and end of the *pepA* gene is integrated into the *E. coli* genome, resulting in the complete replacement of *pepA* with the landing pad cassette, illustrated here in its genomic location between genes *yigP* and *holC*.

E. coli DS941 cells were transformed with λ Red recombineering plasmid pKOBEG-A (Chaverocche *et al.*, 2000), expressing λ Red proteins (Gam, a nuclease inhibitor; Exo, a 5' to 3' exonuclease; and Beta, which protects and promotes annealing to the ssDNA created by Exo) under control of the pBAD promoter which is induced by the addition of arabinose. Successful transformants were isolated by selecting for ampicillin resistance carried by the pKOBEG-A vector. This plasmid is designed to be temperature-sensitive in order to allow removal following successful recombineering; this occurs due to the presence of a mutation in the pSC101 origin backbone which prevents replication of the plasmid at temperatures above 37°C (Hashimoto-Gotohi & Sekiguchi, 1977). Therefore all growth steps consequent to transformation were conducted at 30°C in order to maintain the pKOBEG-A plasmid in the cells. These DS941 cells containing pKOBEG-A were then ready for induction of Red functions and transformation with the LP recombineering PCR product.

Primers were designed to amplify the *EmR* gene with ϕ C31 integrase *attB^{TT}* and *attB^{TC}* sites flanking up- and downstream respectively, along with 50 bp homology sequences to the start and end of the *pepA* gene. These primer sequences are shown in Methods Section 2.4. Using these primers, the recombineering cassette (LPC31) was amplified using a Phusion polymerase protocol nicknamed 'Phussean' (Methods Section 2.13.), which has been optimised for performance with larger primers such as the oligonucleotide primers synthesised for this PCR (123 bp forward primer, 129 bp reverse primer; see Methods Section 2.4). The PCR template used was the *EmR* gene found on the genome of the *E. coli* DS941-derived strain π 1 (see Methods Section 2.5), a strain previously engineered by this lab, in a method that allows amplification directly from the genome by resuspending a colony in water and adding it to the reaction set up (Methods Section 2.13.) Purification of the PCR product was done in one of two ways during recombineering experiments.

At first, PCR reactions were directly run in their entirety on a 1% agarose gel and the correctly sized DNA band (1253 bp) was excised and extracted using a QIAgen Gel Extraction kit. However, after several unsuccessful attempts at recombineering with this PCR product sample, it was thought that exposure to UV during excision of the agarose gel band was having a damaging effect on the DNA, thus decreasing efficiency of the homologous recombineering process. Subsequent PCR products of the LP recombineering fragments were purified using a QIAgen PCR Purification kit immediately after completion of the PCR, and a sample of this eluted DNA was run on a 1% agarose gel to confirm the presence and size of the DNA. This avoided damage to the DNA by UV radiation and increased chances of successful recombineering in subsequent protocols.

The ϕ C31 LP recombineering PCR product was introduced by electroporation into DS941/pKOBEG-A cells that had been induced by L-arabinose for λ Red functions. Following transformation and recovery at 30°C, successful *EmR* LP recombinants were selected by their resistance to erythromycin. Initially the efficiency of recombineering was seen to be very low, with many repetitions of the protocol resulting in no observed *EmR*-positive colonies on erythromycin LB agar plates.

After many iterations of this method, when successful erythromycin-resistant colonies were finally seen (healthy colonies on LB + erythromycin plates following electrotransformation) colony counts were very low, with a maximum of 4 colonies per plate observed. It should be noted that of these successful plates, it was found that incubating agar plates at 37°C rather than 30°C resulted in an improved rate of recombineering. This had previously been suggested to be an improved temperature for the activity of the λ Red proteins, and while the higher temperature decreases the ability of pKOBEG-A to proliferate, sufficient expression of the λ Red proteins should be completed at this stage and maintenance of the plasmid is not essential.

In order to clear transformed cells of the pKOBEG-A temperature-sensitive plasmid, erythromycin-resistant colonies were streaked onto LB agar plates plus erythromycin and grown at 42°C. At this temperature, replication of the pKOBEG-A plasmid is inhibited and the vector does not proliferate. After several generations the temperature-sensitive plasmids should therefore eventually be lost from the cells. After two rounds of re-streaking and growth at 42°C, cells were tested for loss of the pKOBEG-A plasmid by streaking onto LB agar plus ampicillin. Absence of growth on these plates indicated the successful loss of the plasmid.

3.2.3. Confirmation of LP insertion by PCR

To confirm insertion of the ϕ C31 LP cassette into the *pepA* locus, PCRs were designed to read across the flanks of the LP and into the upstream and downstream genes flanking the *pepA* gene. PCR primers were designed to anneal to 20-bp sequences upstream and downstream of *pepA*, in the genes *yjgP* and *holC*, respectively. Forward and reverse primers were also designed to anneal in both the *pepA* gene and *EmR* gene, to be used to test for presence of either gene in the LP location; the former indicating unsuccessful recombineering, the latter indicating successful integration of the ϕ C31 LP cassette. A list of these primer sequences can be seen in Methods Section 2.4.

A diagram representing the five confirmatory PCRs is shown in Figure 3.4. For these reactions, *E. coli* DS941 was used as a negative control. In the case of successful integration of the ϕ C31 LP cassette, PCR1 (upstream to downstream *pepA*), PCR4 (upstream to *EmR*) and PCR5 (*EmR* to downstream) should amplify. In the case of no *pepA* replacement, PCR1, PCR2 (upstream to *pepA*) and PCR3 (*pepA* to downstream) should amplify. PCR1, which amplifies across the whole *pepA* site, should produce a differently sized band dependent on whether the LP integration is successful or unsuccessful, as indicated in the product size table of Figure 3.4.

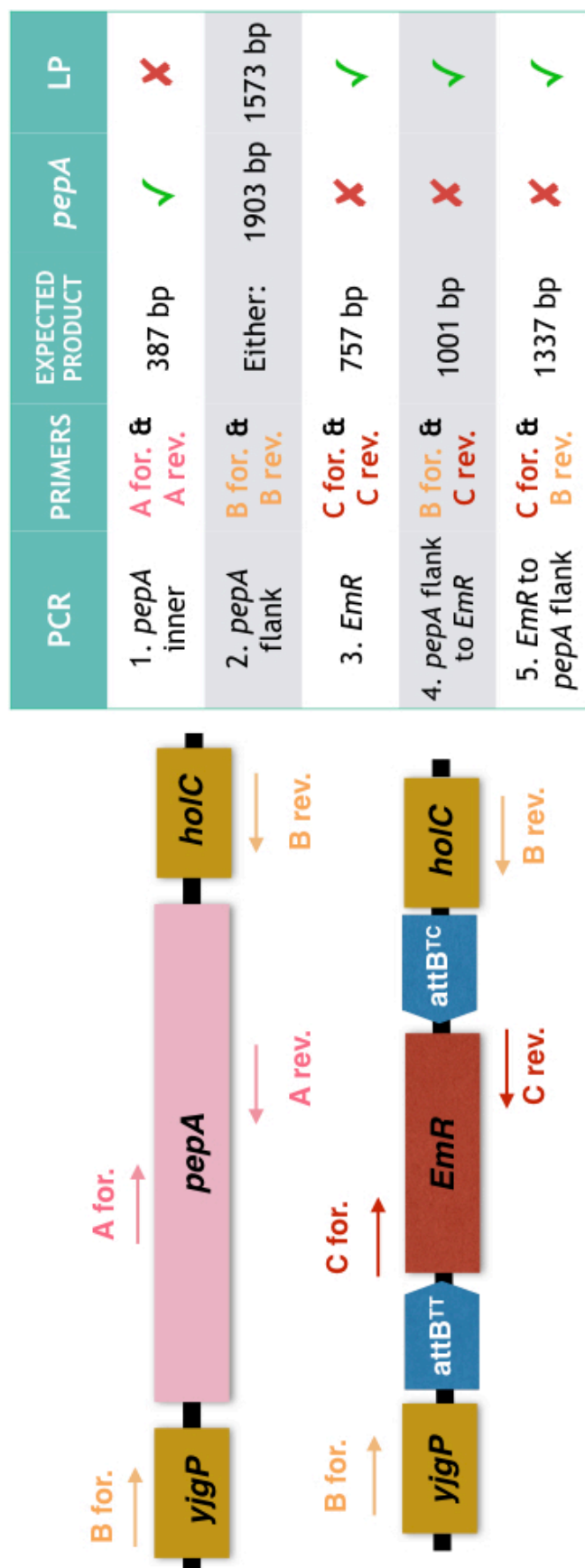


Figure 3.4 : Diagram showing PCR primers used for confirming replacement of the *pepA* gene by the *EmR* ϕ C31 LP; expected sizes of each PCR product are listed, alongside the primers used. PCRs 4 and 5 indicate successful integration of the LP by amplifying across junctions from upstream and downstream of the *pepA* locus into the LP erythromycin resistance marker.

These five confirmatory PCRs were performed on a selected erythromycin-resistant colony, obtained from a recombineering transformation, alongside DS941 as a negative control. Figure 3.5 shows these PCRs run on a 1% agarose gel with a 1 kb marker ladder. According to the expected band sizes, the erythromycin-resistant recombineering transformant, from here on labelled as LPC311a, is positive for *pepA* replacement with LPC31. Particularly convincing is the amplification of PCR4 and PCR5, which show amplification from the *EmR* gene to the flanking regions of *pepA*. In addition, PCR2 shows amplification of the different expected bands from the original DS941 and LPC311a, with no amplification of *pepA* in the latter, suggesting complete replacement by recombineering. Strain LPC311a was grown overnight in LB liquid culture and a glycerol stock was created to preserve the successful LP strain for future use.

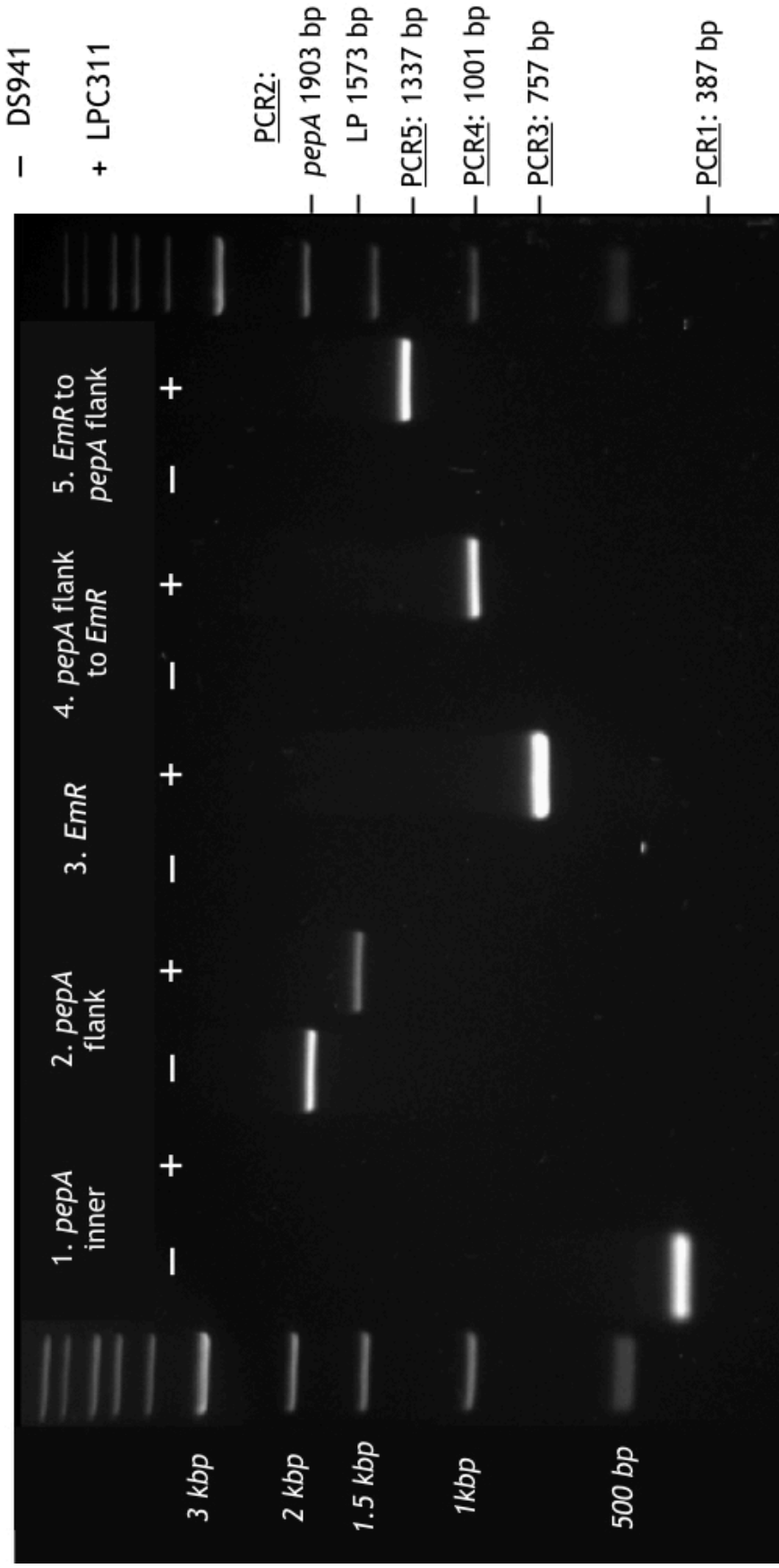


Figure 3.5: Agarose gel of LPC311 *pepA* replacement confirmation PCRs as diagrammed in Figure 3.4 performed on strains DS941 (-) and LPC311a (+). PCR1 only amplifies *pepA* from DS941 cells as expected; PCR2 shows amplification of products of different sizes in DS941 (1903 bp) and LPC311a (1573 bp) as expected; PCR3 only amplifies *EmR* in LPC311a cells, confirming presence of the marker; PCR4 and PCR5, which amplify from the LP to flanking sequences of *pepA*, both show expected band sizes (1001 bp and 1337 bp respectively), strongly suggesting that the LP cassette has been successfully inserted into the *pepA* locus.

3.3. Development of an appropriate method for cassette delivery

3.3.1. Designing a temperature-sensitive donor vector

The initial ‘donor’ cassettes to be integrated into LPC31 consisted of an antibiotic resistance marker for either chloramphenicol (*CmR*) or kanamycin (*KmR*) flanked by the *attP^{TT}* and *attP^{TC}* sites which correspond to the LPC31 *attB* sites. These donor cassettes, when transformed into LPC311 cells expressing ϕ C31 integrase, were designed to recombine with their corresponding *attB* sites resulting in cassette exchange across the LP, replacing the *EmR* gene with *CmR* or *KmR*. A diagram showing these donor cassette designs and the recombination reactions by which cassette exchange occurs across the LP is shown in Figure 3.6.

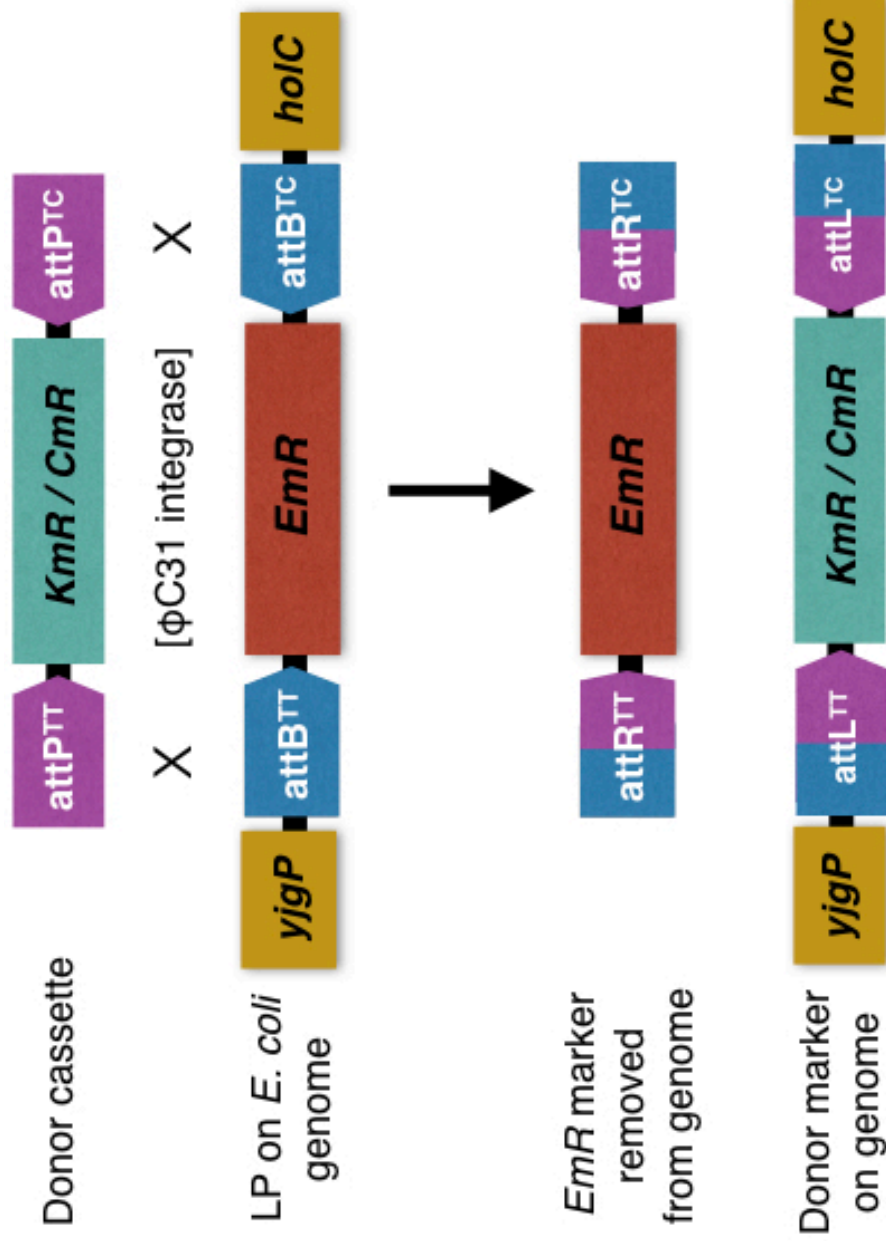


Figure 3.6 : Diagram of the designed phiC31 donor cassette, consisting of an antibiotic resistance marker gene (for kanamycin, *KmR*, or chloramphenicol, *CmR*), flanked by phiC31 integrase attachment sites *attP^{TT}* and *attP^{TC}*. When introduced into the LPC311 strain expressing phiC31 integrase, *attP^{TT}* recombines with *attB^{TT}*, and *attP^{TC}* with *attB^{TC}*, to facilitate cassette exchange, replacing the *EmR* LP marker with the donor marker. The donor marker then remains on the genome flanked by the two recombination product sites *attL^{TT}* and *attL^{TC}*, while the *EmR* marker is removed alongside the two other product sites *attR^{TT}* and *attR^{TC}*.

To begin with, donor cassettes were to be delivered via a plasmid vector. A desirable aim of the LP system is to eventually be able to transform the donor cassette as a PCR product. However, using a plasmid vector first would allow for stable transformation of the cassette and for LP integration to be demonstrated before attempts with linear DNA (which faces the problem of degradation by native *E. coli* endonucleases following transformation).

Using a plasmid vector for donor cassette delivery presented the problem of marker detection following cassette exchange, as the donor plasmid, which contains the marker before integration, can also express the antibiotic resistance gene (for chloramphenicol or kanamycin). The solution to this was to use a temperature-sensitive plasmid vector, which could be removed from cells following cassette exchange by growing transformants at a higher temperature. Once cells are cleared of the donor plasmid they can be tested for resistance to the donor marker antibiotic to confirm cassette exchange into the genome.

The temperature-sensitive vector chosen was based on the plasmids used for λ Red recombineering transformations, pKOBEG-A (*ApR* backbone), or alternatively pKOBEG-C (*CmR* backbone) (Chaverocche *et al.*, 2000). This plasmid contains the *repA101ts* gene which encodes the protein RepA101ts, and the pSC101 origin of replication. RepA101ts binds to the origin to initiate replication at a temperature of 30°C, but becomes inactive at temperatures between 37°C and 42°C (Hashimoto-Gotohi & Sekiguchi, 1977). Therefore, when cells containing the temperature-sensitive plasmid are grown at 42°C, the plasmid is no longer able to replicate and is lost. This function is ideal for temporary delivery of the donor cassette.

Donor cassettes for *CmR* and *KmR*, as shown in Figure 3.6, were produced via PCR using primers containing the desired *attP* sequences that amplified the chloramphenicol resistance gene (*cat*) and kanamycin resistance gene (*aph*) from the template plasmids pSW23 and pSW29 (Demarre *et al*, 2005), respectively. Included were restriction endonuclease sites for *HindIII* (forward *attP^{TT}*) and *KpnI* (reverse *attP^{TC}*) for cloning into pKOBEG-A or pKOBEG-C, replacing completely the λ Red protein genes and pBAD promoter. These primers can be seen in Methods Section 2.4.

Initial cassette vector cloning yielded three combinations of donor marker and vector backbone resistance genes as follows: pRM1 (*CmR* donor, *ApR* backbone), pRM2 (*KmR* donor, *CmR* backbone), pRM3 (*KmR* donor, *ApR* backbone). These plasmids are shown in Methods Section 2.5.

3.3.2. Development of the ϕ C31 LP cassette exchange protocol (temperature-sensitive donor plasmid)

ϕ C31 LP cassette exchange transformations were based primarily on transformations used for recombineering. This used a protocol for high efficiency transformation of electrocompetent cells, and using a similar method to recombineering has the added advantage of providing a comparison of efficiency between the homologous recombination approach and the use of serine integrase enzymes.

ϕ C31 integrase was first expressed constitutively in cells from the plasmid pFEM32, which carries ampicillin resistance (Methods Section 2.5). pFEM32 was introduced into LPC311a cells first by chemical transformation. LPC311a/pFEM32 cells were subsequently transformed with pRM2 (*KmR* donor, *CmR* backbone) by LP cassette exchange electrotransformation (see Methods Section 2.17). Following recovery, successful transformants were selected on LB agar plates containing kanamycin at 30°C. At this temperature, pRM2 should be maintained inside the LPC311a cells.

In order to then clear LPC311a cells of pRM2 and measure for cassette exchange, 100 colonies were taken from LB-kanamycin transformation plates and patched by toothpick onto another LB agar plates with no antibiotic. These 100-grid plates were then incubated at 42°C overnight. This process was repeated the next day onto another 100-grid LB agar plate (no antibiotics) to be incubated at 42°C for a second overnight growth. The reason for not including kanamycin in these LB plates at this stage was to avoid selecting for any remaining pRM2 expressing the *KmR* gene while in the process of trying to remove the plasmid from the cells by heating. This had been a problem in initial cassette exchange experiments, where selecting for cassette integration by kanamycin resistance seemed to override the temperature-based removal of the donor plasmid due to the backbone also carrying the *KmR* gene.

After two 42°C overnight incubations, each of the 100 transformant patches were transferred from the LB plates containing no antibiotics to three 100-grid LB plates containing either erythromycin (to test for remaining unreplaced *EmR* LP), chloramphenicol (to test whether cells had been successfully cleared of pRM2) or kanamycin (to test for successful cassette exchange in the case that pRM2 had been removed from the cells).

In each set of 100 transformants tested across these plates, it was found that cells were still erythromycin resistant but had lost the pRM2 plasmid, with each patch growing on LB plus erythromycin but not on LB plus kanamycin, indicating that cassette exchange had been unsuccessful and the erythromycin resistance gene remained in place at the LP site. This suggests that cassette exchange has been unsuccessful in all pRM2 transformants and there appeared to be an issue with the cassette exchange protocol or the LP cassette sequence.

3.4. Addressing problems with ϕ C31 integrase expression in the LP strain LPC311

3.4.1. Measuring ϕ C31 integrase activity by the GalK recombination assay

To determine the cause of the unsuccessful cassette exchange, the activity of ϕ C31 integrase in the LPC311a cells was investigated. This was done using a GalK recombination (MacConkey) assay. The assay is an *in vivo* test for recombinase activity, which uses a substrate plasmid containing *attB* and *attP* sites in a head-to-tail orientation around a reporter gene encoding galactokinase, *galK*, which is required for the metabolism of galactose.

When cells containing the complete substrate expressing *galK* are grown on plates of MacConkey agar containing galactose as the only carbon source and a pH indicator (2-methyl-3-amino-6-dimethylaminophenazine) which shows red at pH < 6.8 and yellow at pH > 8.0, colonies appear red or pink. If recombination occurs between *attB* and *attP*, the *galK* gene is lost and cells are unable to metabolise galactose, forcing them to metabolise amino acids in the MacConkey agar, raising the pH and resulting in the formation of white colonies.

Following visual confirmation of red or white colony formation on MacConkey agar, further confirmation of recombination can be observed by extracting the plasmid DNA from the cells. This plasmid DNA is then cut with a restriction endonuclease (*NruI*, for which there are two sites on the substrate plasmid), then running the resulting DNA on an agarose gel to check the size of the fragments. An illustration of this assay and the sizes expected for substrate pGD001 are shown in Figure 3.7.

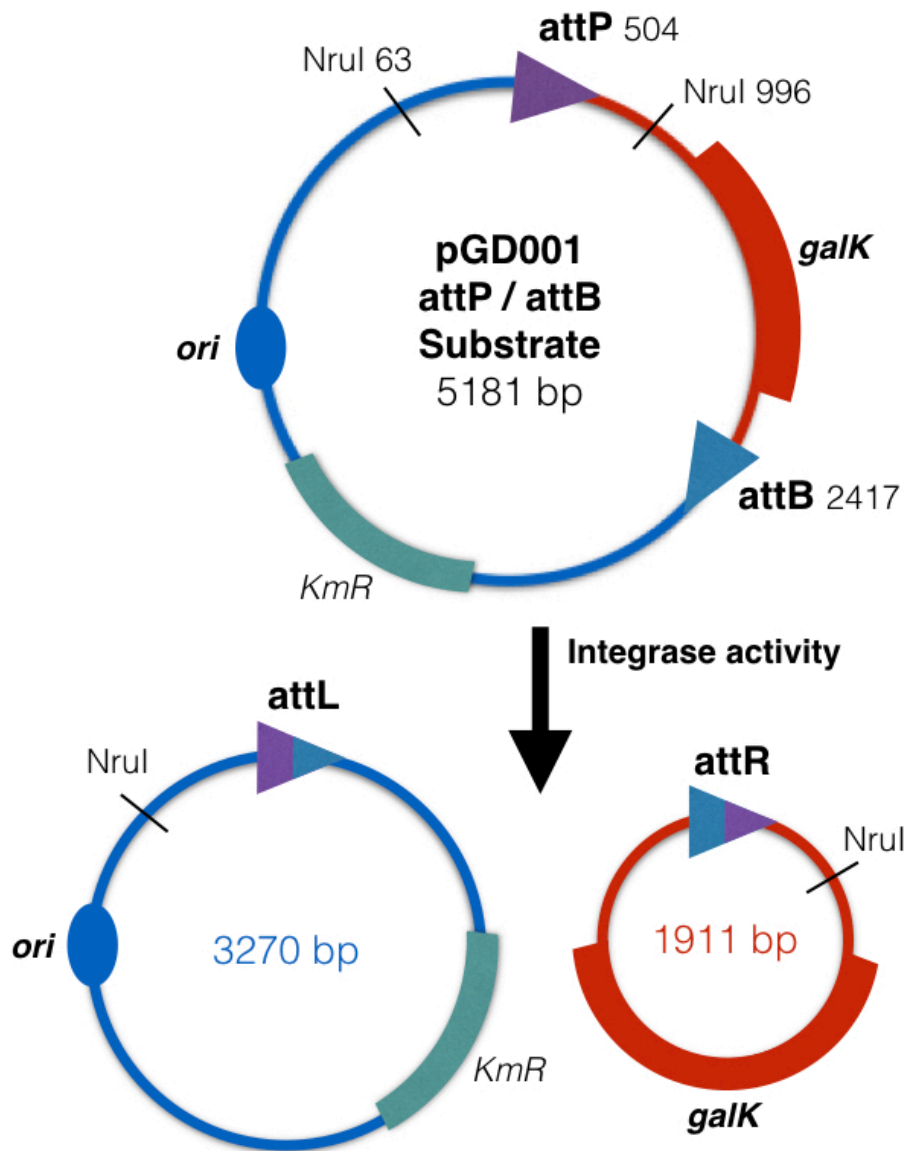


Figure 3.7: Diagram illustrating MacConkey recombination assay used to measure ϕ C31 integrase activity in DS941 and LPC311a cells. When the substrate plasmid, pGD001 (see Methods 2.5), is introduced by electrotransformation into *E. coli* cells expressing ϕ C31 integrase, recombination should occur between attachment sites *attP* and *attB*, resulting in the excision of a DNA fragment containing the *galk* gene and the formation of two plasmids, the larger of which consists of the plasmid backbone containing the origin of replication and a smaller plasmid containing *galk* but no origin; the new smaller *galk*-containing plasmid is therefore lost from the cells due to its lack of an origin, while the larger plasmid is maintained by selecting for kanamycin resistance. When cut with restriction endonuclease *NruI*, the substrate produces two fragments of sizes 933 bp and 4248 bp, while recombined plasmid produces a single 3270 bp fragment.

DS941 and LPC311a cells containing pFEM32 (which constitutively expresses ϕ C31 integrase) were transformed via a standard electroporation protocol with the ϕ C31 *attB-attP galK* substrate plasmid pGD001. DS941 and LPC311a cells containing pFEM35 (constitutive expression of Bxb1 integrase) were also transformed as a recombination-negative control. Recovered cells were then spread on MacConkey agar plates containing galactose and kanamycin (pFEM32 and pFEM35 vector selection) and incubated overnight at 37°C. Photographs of these plates are shown in Figure 3.8.

MacConkey transformation plates were then scraped (1 mL liquid LB added and plate swept with a sterile glass rod to dislodge cells from colonies) to collect as many cells as possible. The plasmid DNA was then extracted with a QIAgen Plasmid DNA Miniprep Kit and digested with *NruI* to be run on an agarose gel. This agarose gel is shown in Figure 3.9.

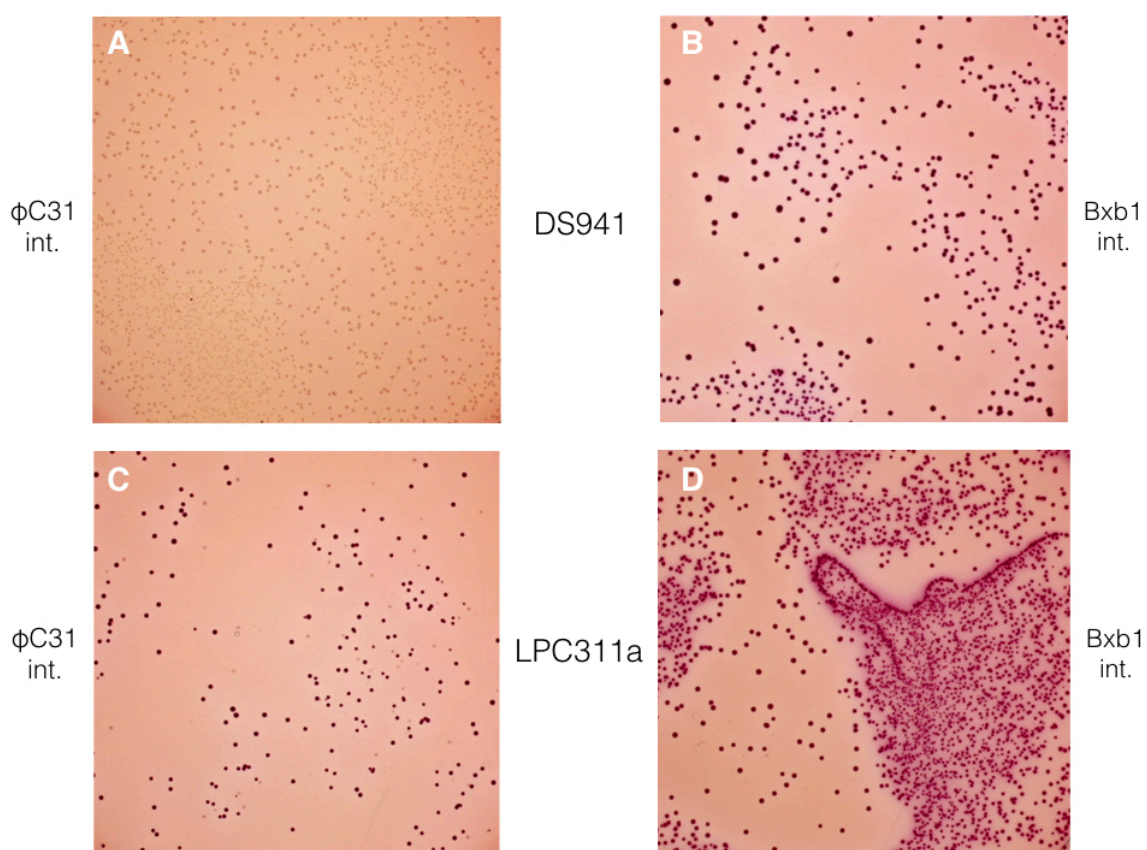


Figure 3.8: *Galk* assay MacConkey plates showing transformations of substrate plasmid pGD001: A) DS941/pFEM32 expressing ϕ C31 integrase, B) DS941/pFEM35 expressing Bxb1 integrase, C) LPC311a/pFEM32 expressing ϕ C31 integrase, D) LPC311a/pFEM35 expressing Bxb1 integrase. The high percentage of white or cream coloured colonies seen in A) indicate efficient recombination. By comparison, a much lower number of white colonies are seen in C), suggesting that ϕ C31 integrase recombination activity is much lower in LPC311a/pFEM32 cells compared to DS941/pFEM32 cells.

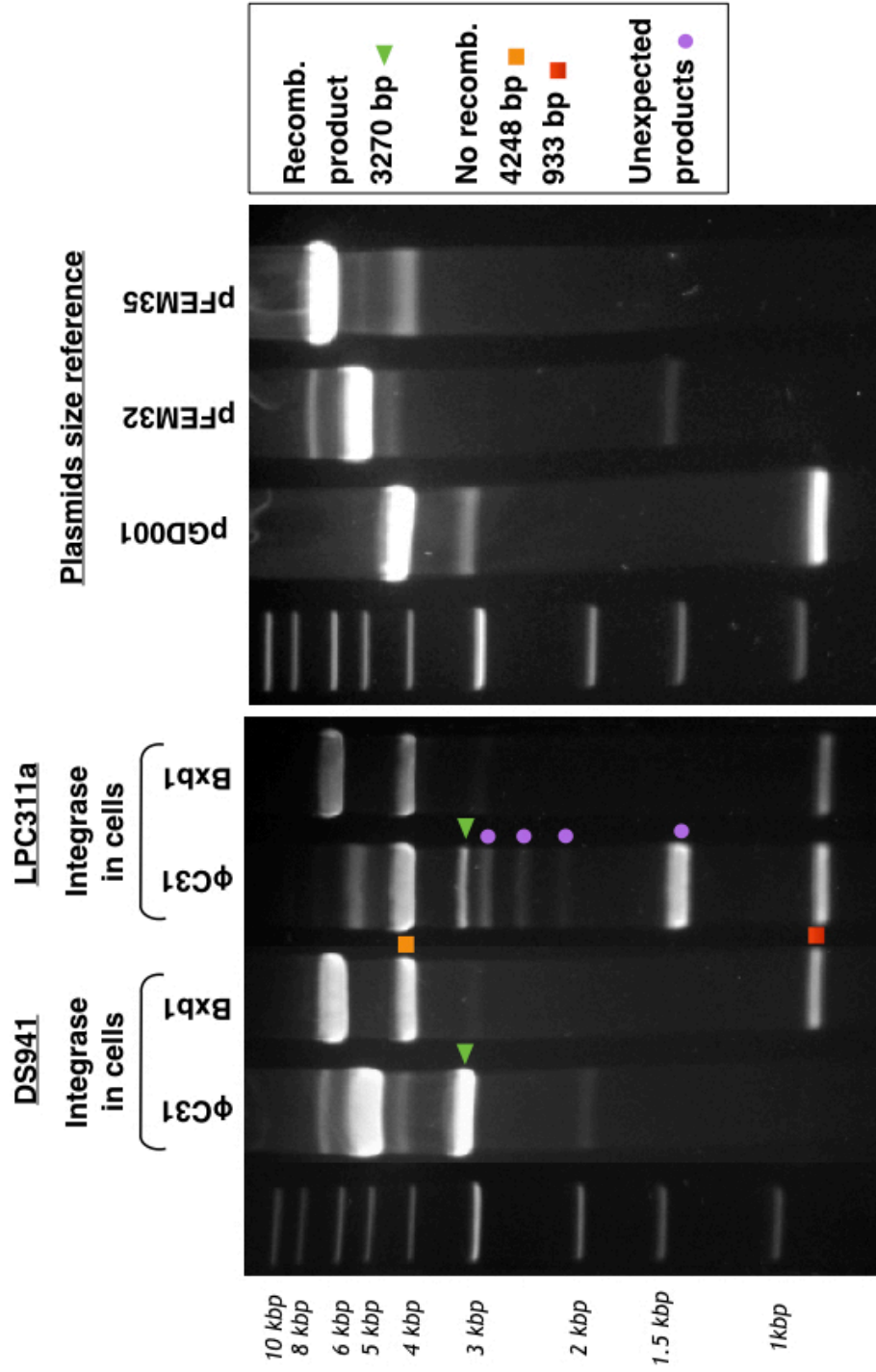


Figure 3.9: Agarose gel showing DNA extracted from each of four *GalK* assay plates, containing recombined or unrecombined substrate plasmid pGD001 cut with restriction enzyme *NruI*, and either pFEM32 (ϕ C31 integrase) or pFEM35 (Bxb1 integrase). Recombination product of 3270 bp is indicated by a green arrow head; unrecombined products are indicated with a red and orange squares; 'unexpected products' observed in the gel are highlighted with purple circles. An NEB 1kb ladder is included for size reference.

In Figure 3.9, DS941 cells expressing ϕ C31 integrase show almost complete recombination of the pGD001 substrate, producing a strong band of 3270 bp in length. In contrast, and in line with results observed from MacConkey plates, LPC311a cells expressing ϕ C31 integrase show very little recombination, leaving strong bands of unrecombined pGD001 (cut with *Nru*I) of lengths 4248 bp and 933 bp. Unexpected bands were also observed in this DNA sample. The appearance of these bands, indicated in Figure 3.9 by purple circles, appears to be in balance with the disappearance of the pFEM32 band in this sample (the highest band in this lane).

3.4.2. Determining the fate of the ϕ C31 integrase expression plasmid in LPC311a cells

After detecting unexpected bands in the LPC311a *galk* assay DNA samples, suspected to be connected to the loss of ϕ C31 integrase expression plasmid pFEM32 as seen in the agarose gel shown in Figure 3.9, the next step was to investigate what was happening to this plasmid during ϕ C31 integrase expression in LPC311a cells. The unexpected DNA bands were not seen in DS941 cells, which do not feature the inserted ϕ C31 LP cassette containing two *attB* sites, indicating that these sites may be having an effect on the expression plasmid and possibly the health of the *E. coli* cells.

To begin investigating the effect of continuous ϕ C31 integrase expression from the pFEM32 plasmid in LPC311a cells, DS941 and LPC311a glycerol stocks containing either pFEM32 or pFEM35 (Bxb1 integrase) were streaked onto LB agar plates containing ampicillin to select for the *ApR*-carrying plasmids, and erythromycin for LPC311a only to select for the LP *EmR* marker. It was observed that LPC311a cells containing pFEM32 (expressing ϕ C31 integrase) grew far less on these agar plates than LPC311a cells containing pFEM35, or DS941 cells containing either plasmid. Photographs of these plates are shown in Figure 3.10.

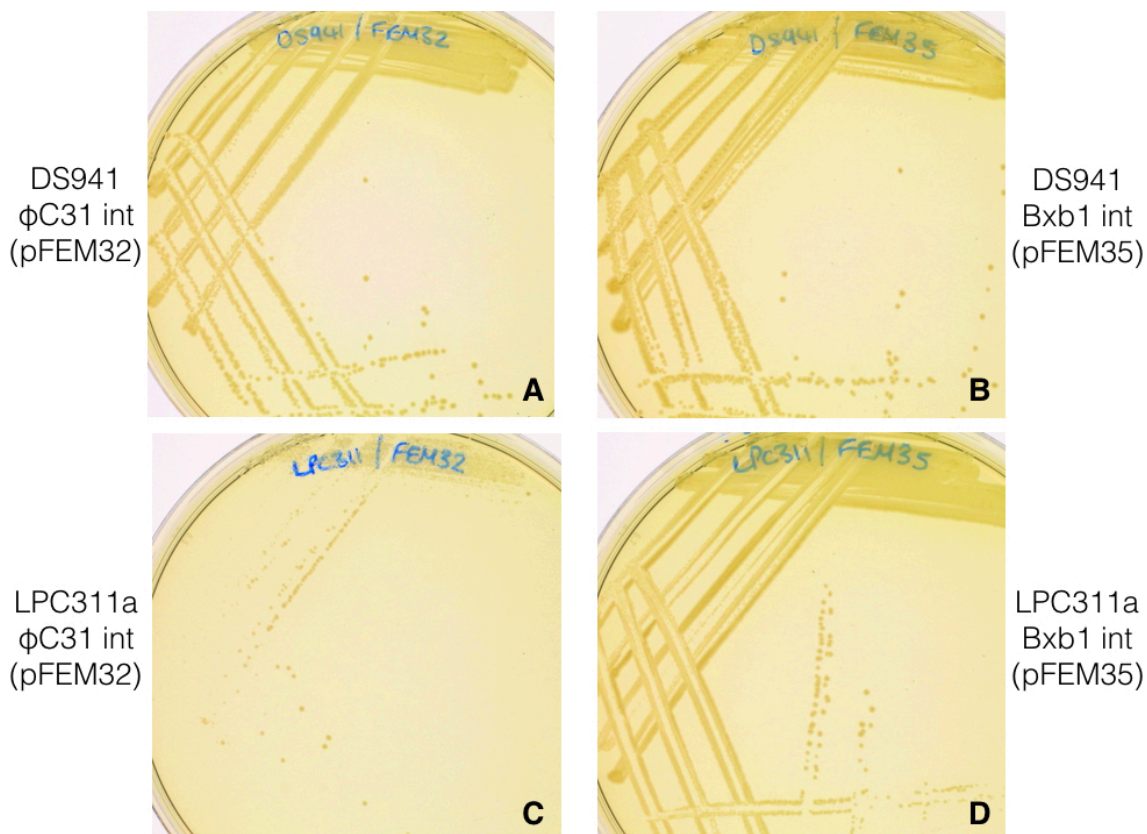


Figure 3.10: *E. coli* cells containing constitutive expression plasmids for φC31 integrase (pFEM32) and Bxb1 integrase (pFEM35) control: A) DS941+pFEM32, B) DS941+pFEM35, C) LPC311a+pFEM32, D) LPC311a+pFEM35. Note the poor growth of LPC311a cells expressing φC31 integrase on plate C. Cells are streaked onto LB agar containing ampicillin to select for the *Ap^R*-carrying expression plasmid (A, B, C & D) and erythromycin to select for *Em^R* on the LPC311 genome (C & D only).

While this could be seen as a possible indication that ϕ C31 integrase expression inhibits growth of LPC311a cells, it is also important to note that because these streaks were taken by toothpick from a glycerol stock, the number of cells streaked onto the plates pictured could have been inconsistent between the plates. Nonetheless, it was a consistent observation that LPC311a/pFEM32 cells grew less on ampicillin and erythromycin plates, as also seen in the low colony count of LPC311a/pFEM32 MacConkey assay plates in Figure 3.8.

To find out what was happening to the pFEM32 plasmid in LPC311a cells, several single colonies were taken and grown overnight in liquid LB media plus ampicillin and erythromycin. Additionally, a single colony of DS941/pFEM32 was also selected and grown overnight with ampicillin selection. Plasmid DNA was then extracted from these overnight cultures (using a QIAgen Plasmid DNA Miniprep Kit), and this DNA was run on an agarose gel. Figure 3.11 shows an agarose gel of pFEM32 extracted from DS941 and five LPC311a colonies. In the agarose gel shown here, the DNA is uncut and therefore the brightest bands observed are supercoiled DNA, with an above fainter nicked DNA band (better indicating the size of the plasmid).

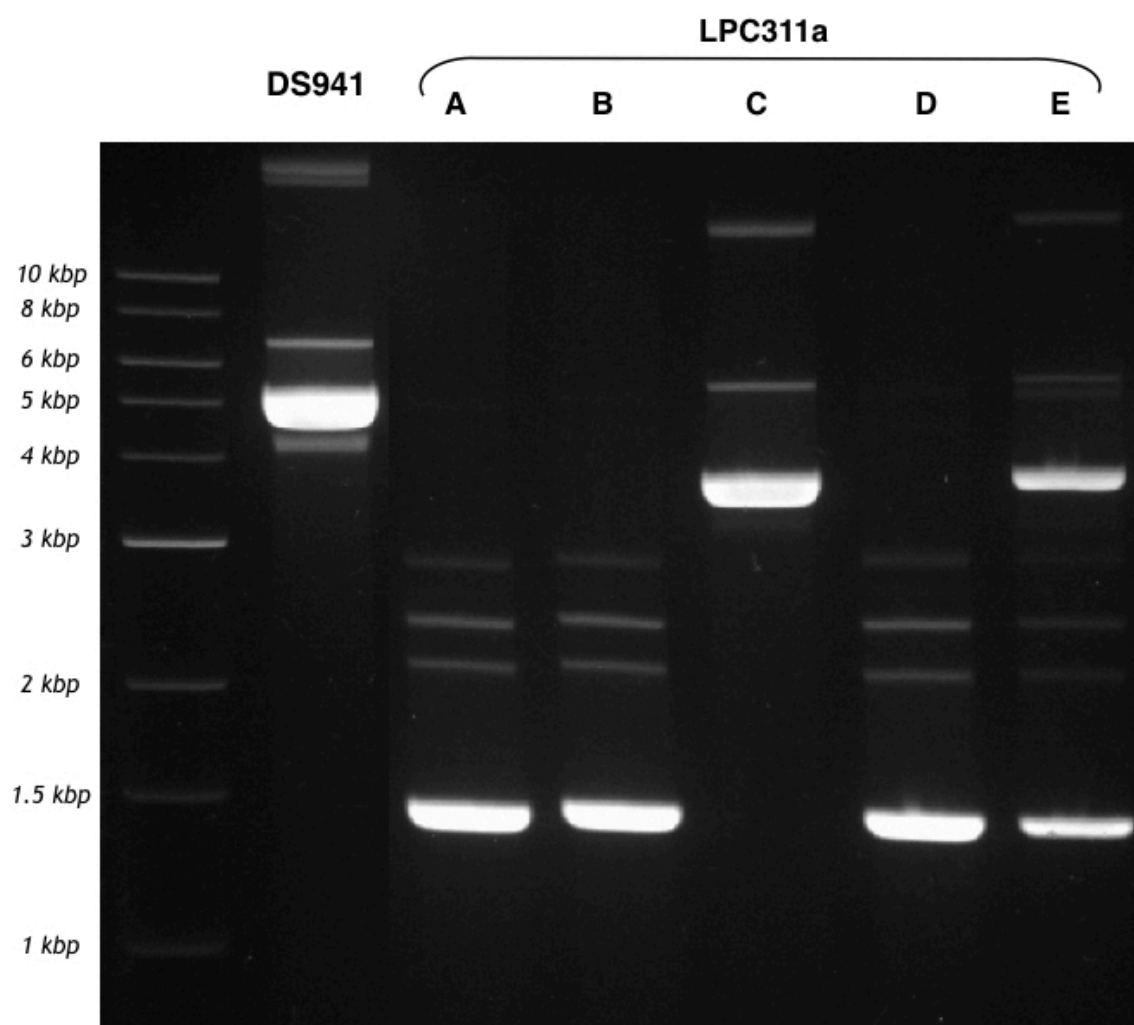


Figure 3.11: pFEM32 extracted from DS941 cells and 5 single LPC311a colonies, A-E, run on a 1% agarose gel by electrophoresis. NEB 1kb ladder included for size reference.

In LPC311a, pFEM32 appears to be converted to two different fragments of DNA at sizes ~ 5 kbp and ~ 2.2 kbp. In each of these five colonies, the plasmid appears to have been completely converted to either one or both of these plasmids. It can be assumed that these products are replicating plasmids and still contain the origin of replication and selection marker as they were maintained *in vivo*, and would have been lost if linear. In contrast, as previously indicated, pFEM32 is stable in DS941 and there is no conversion to these products.

The cause of this plasmid instability could be due to an interaction between ϕ C31 integrase and the LP *attB* sites. One possibility is that due to constant expression (and possible build-up) of integrase, the enzyme might be recombining one of these LP sites with a pseudo-*attP* site in the *E. coli* genome, with a negative effect on the viability of the cells. This would explain the poor growth of LPC311a + pFEM32 on agar plates. In these cells, the reaction against the negative events caused by ϕ C31 integrase activity appears to be deletion of the expression plasmid, acting as a counter-selection against ϕ C31 integrase and resulting in poor recombination activity within the cells. For the continuation of this project it was therefore clear that constitutive expression of ϕ C31 integrase in LPC311a cells causes genetic instability, and an alternative expression solution must be implemented.

3.4.3. An inducible ϕ C31 integrase expression method for improved activity in LPC311a

To overcome the problems observed with continuous expression of ϕ C31 integrase, a decision was made to regulate ϕ C31 integrase expression more tightly using an inducible promoter. A plasmid vector which expresses ϕ C31 integrase under control of the arabinose-inducible pBAD promoter complex, pZJ7, had already been cloned and used in this lab (Methods Section 2.5). The pBAD promoter is induced by addition of L-arabinose and inhibited by the addition of glucose (Guzman *et al.*, 1995).

The pZJ7 ϕ C31 integrase inducible expression plasmid (*CmR*) was introduced into DS941 and LPC311a cells by chemical transformation (Methods Section 2.9.1), with transformants selected by chloramphenicol resistance. *In vivo* recombination assay transformations with the *Galk attP-attB* substrate pGD001 were repeated as previously (Methods Section 2.21) with two alterations. Firstly, in order to suppress expression of ϕ C31 integrase, glucose was added to the overnight starter cultures before inoculation into larger cultures for making electrocompetent cells. Secondly, arabinose was added to induce ϕ C31 integrase expression under control of the pBAD promoter half way through growth of electrocompetent cell cultures (~1.5 hours uninduced growth followed by ~1.5 hours induction). A control culture was also set up containing glucose throughout growth in order to suppress expression of ϕ C31 integrase. Cells recovered after electroporation were spread onto MacConkey plates containing galactose and kanamycin (selecting for pGD001). Photographs of these plates are shown in Figure 3.12.

To confirm whether inducible ϕ C31 integrase expression leads to more efficient recombination in LPC311a cells compared to previous constitutive expression, MacConkey plates were scraped and pGD001 DNA was extracted. This DNA was then run, non-linearised, on an agarose gel. This gel is shown in Figure 3.13, where a clear band of recombined pGD001 (3270 bp, supercoiled) can be seen only in pZJ7-containing cells induced with arabinose (lane 'ara +'). While there is still a strong band of unrecombined pGD001, indicating incomplete recombination, this is still a significant level of recombination and is in sharp contrast to cells with glucose suppression rather than arabinose induction (lane 'ara -, gluc +'), where no recombination product can be seen.

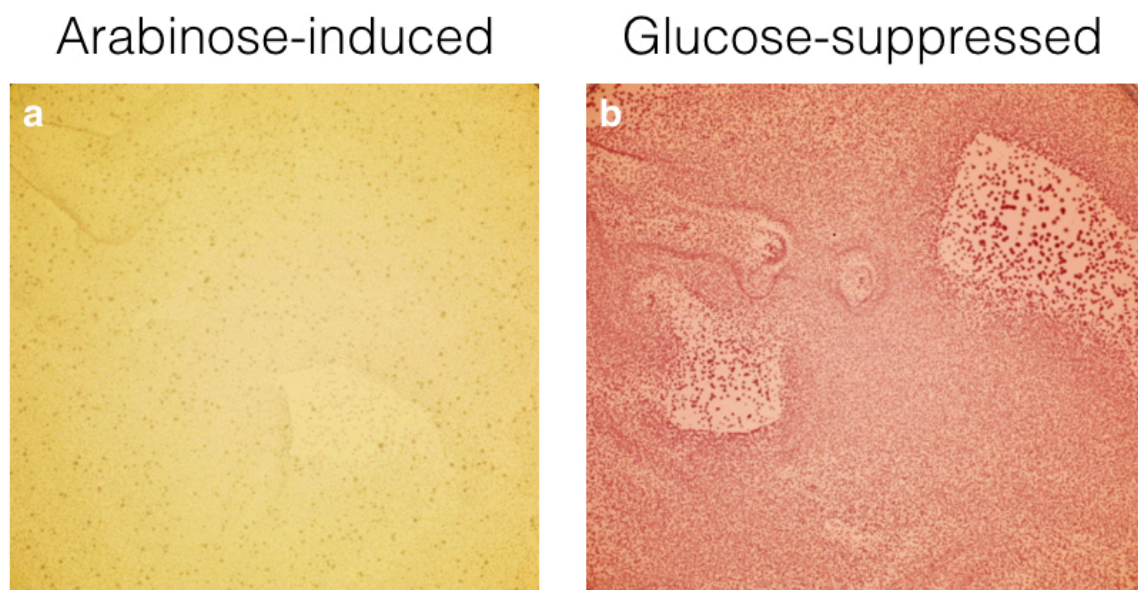


Figure 3.12: MacConkey agar plates from *Galk* recombination assay pGD001 transformations into a) LPC311a cells containing arabinose-induced pZJ7 (therefore expressing ϕ C31 integrase) and b) LPC311a cells containing pZJ7 suppressed by addition of glucose. Complete recombination is indicated by presence of only yellow/pale-coloured colonies in the arabinose-induced cells (a). In contrast, the glucose-suppressed cells (b) only give red colonies, indicating incomplete (or zero) recombination between *attB* and *attP*.

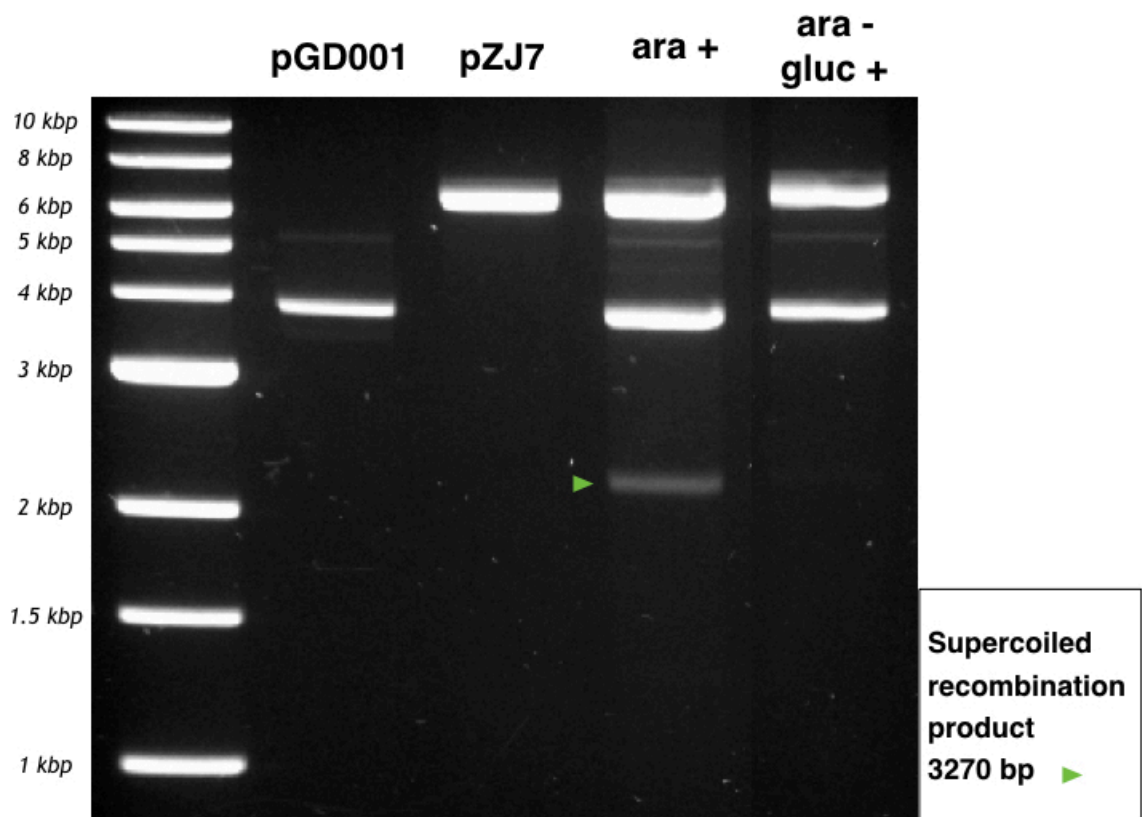


Figure 3.13: Agarose gel showing DNA extracted from MacConkey plates of LPC311a cells containing pZJ7, transformed with pGD001 after being grown under two conditions: 1) ara+, 3-hour induction with 10% L-arabinose; 2) ara-gluc+, no arabinose induction, but with the addition of 10% glucose to suppress ϕ C31 integrase expression. pGD001 (5181 bp) and pZJ7 (7175 bp) plasmids are run alongside as a reference. The expected recombination product is a 3270 bp plasmid, indicated here with a green arrow. Due to the circular supercoiled nature of this plasmid (samples were not linearised with *Nru*I following extraction), the recombination product size does not match size references of the linear reference ladder, which is shown for comparison.

3.5. Optimisation of ϕ C31 integrase LP cassette exchange protocol

3.5.1. Transformation of donor cassette delivery vector into LPC311 cells with induced ϕ C31 integrase expression

With ϕ C31 integrase expression controlled in LPC311a cells by the arabinose-inducible pBAD promoter on plasmid vector pZJ7, LP cassette exchange experiments were repeated. LPC311a cells that had first been transformed with pZJ7 were then transformed with the temperature-sensitive *KmR* LP donor plasmid pRM3 (Methods Section 2.5). This was done by following the previously used high efficiency electrotransformation protocol for recombineering in which arabinose was added for induction prior to cell cultures reaching log growth phase (Methods Section 2.15). Cells recovered after electroporation were plated onto LB agar containing chloramphenicol, selecting for the backbone of pRM3. Selecting for the backbone rather than the donor cassette means that the number of successful plasmid transformants could then be counted. The percentage of these transformants that had successfully performed cassette exchange could then be calculated by testing large numbers of colonies for loss of erythromycin resistance and gain of kanamycin resistance once the cells had been cleared of pRM3 by heating at 42°C. Cassette exchange could then be confirmed beyond doubt using PCRs that read across the *att* site junctions and DNA sequencing of the PCR products.

3.5.2. Detection of cassette exchange

3.5.2.1. Antibiotic-resistance markers

Following transformation of induced LPC311a/pZJ7 cells with *KmR* donor plasmid pRM3, colony counts on the LB agar recovery plates containing chloramphenicol averaged around 1000 colonies (Methods Section 2.17). 100 of these single colonies were taken and patched onto a 100-grid LB agar plate containing no antibiotic, to be incubated overnight at 42 °C. From this overnight plate, the patching process was repeated onto a second LB agar plate (no antibiotic) and again incubated at 42 °C overnight (Methods Section 2.16). Patches were then transferred by toothpick onto three different LB agar plates to test for resistance to three antibiotics: chloramphenicol to detect continued presence of *CmR*-carrying pRM3; kanamycin, which in combination with loss of *CmR* will indicate successful cassette exchange of *KmR* into the LP; and erythromycin to test for cassette exchange by loss or retention of the *EmR* LP marker gene.

Initial antibiotic patch tests revealed that 100/100 colonies remained resistant to all three antibiotics. This was first thought to be an indicator that the temperature-sensitive pRM3 was not being removed from the cells by heating and thus preventing determination of whether cassette exchange was occurring. If the temperature-sensitive plasmid was indeed persisting in the cells, the observation of all three resistances could indicate one of two outcomes; either no cassette exchange occurs and pRM3 remains in the cells unchanged, or cassette exchange takes place leaving kanamycin resistance at the LP site on the genome, with the erythromycin resistance marker *EmR* taking its place on the temperature-sensitive vector plasmid.

Several iterations of the ϕ C31 integrase cassette exchange transformation protocol were tested. Altered protocols featuring induction of ϕ C31 integrase expression with arabinose at different times during growth were explored, as well as suppressing ϕ C31 integrase expression with glucose until the moment of induction, but these all led to the same outcome as described above.

Several methods of recovery and growth following transformation were tested to determine whether temperature had an effect on cassette exchange or temperature-sensitive plasmid behaviour. Firstly, liquid recovery cultures following transformations were tested at 30°C, 37°C or 42°C, as well as incubating post-transformation plates at 30°C, 37°C or 42°C, but no difference was seen in the outcome of the protocol; all transformants tested on 100-grid plates remained resistant to all three antibiotics, indicating that cassette exchange was not taking place successfully.

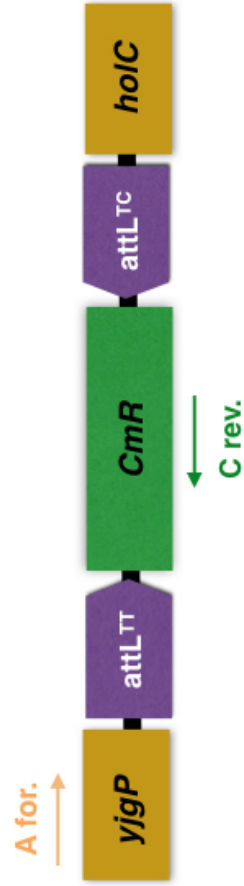
Transformants (kanamycin-resistant colonies) were also subjected to multiple lengths and rounds of heat treatment at 42°C in order to make sure cells were cleared of the temperature-sensitive donor plasmid. When plasmid DNA was extracted from these cells, samples showed no presence of pRM3 or any other similarly sized plasmid. This led to the suspicion that something else was happening in the cells in order for them to remain resistant to kanamycin, chloramphenicol and erythromycin without containing the donor plasmid that carries *KmR* and *CmR*. After this, a series of PCR reactions was devised to determine if any recombination or integration was occurring at the ϕ C31 integrase LP site.

3.5.2.2. PCR amplification of landing pad region

After several attempts to show cassette exchange using gain and loss of antibiotic resistance markers, PCRs were designed to check whether cassette exchange had taken place at the ϕ C31 integrase LP site in LPC311a. Similar to the PCRs used to confirm the integration of the ϕ C31 LP into the *pepA* locus by recombineering (see Section 3.2.3, Figures 3.4 and 3.5), primers were designed to amplify a region from outside the LP and across the *att* site junctions.

In initial cassette exchange experiments, temperature-sensitive donor plasmid pRM3 was used as the donor plasmid, which features *CmR* as the donor cassette marker and an *ApR* backbone. However in later experiments, pRM2 (*KmR* donor, *ApR* backbone) was used to avoid antibiotic resistance marker incompatibility with the new pZJ7 expression plasmid, which also carried a *CmR* marker. pRM2 and pRM3 are detailed further in Methods Section 2.5.

The possible outcomes of initial pRM3 cassette exchange experiments should be either: no integration occurs and *EmR* remains in place, or cassette exchange is successful leaving *CmR* in the LP site flanked by two *attL* sites. The outcomes of initial cassette exchange transformations in LPC311a with pRM2 as a donor were probed by designing a PCR with primers that read from outside *pepA* into *CmR*, and therefore only amplify if *CmR* has been successfully integrated into the LP. A diagram of this PCR is shown in Figure 3.14, and primers used are listed in Methods Section 2.4.



PCR	PRIMERS	EXPECTED PRODUCT
pepA flank to <i>CmR</i>	A for. & C rev.	747 bp

Figure 3.14: Diagram showing the confirmatory PCR used to check many LPC311a / pZJ7 + pRM3 transformants for cassette exchange. Forward primer (A. for) reads from within the *yigP* gene upstream of *pepA*, and the reverse primer reads from within the *CmR* marker gene. The product of this PCR should be 747 bp in length if cassette exchange has occurred.

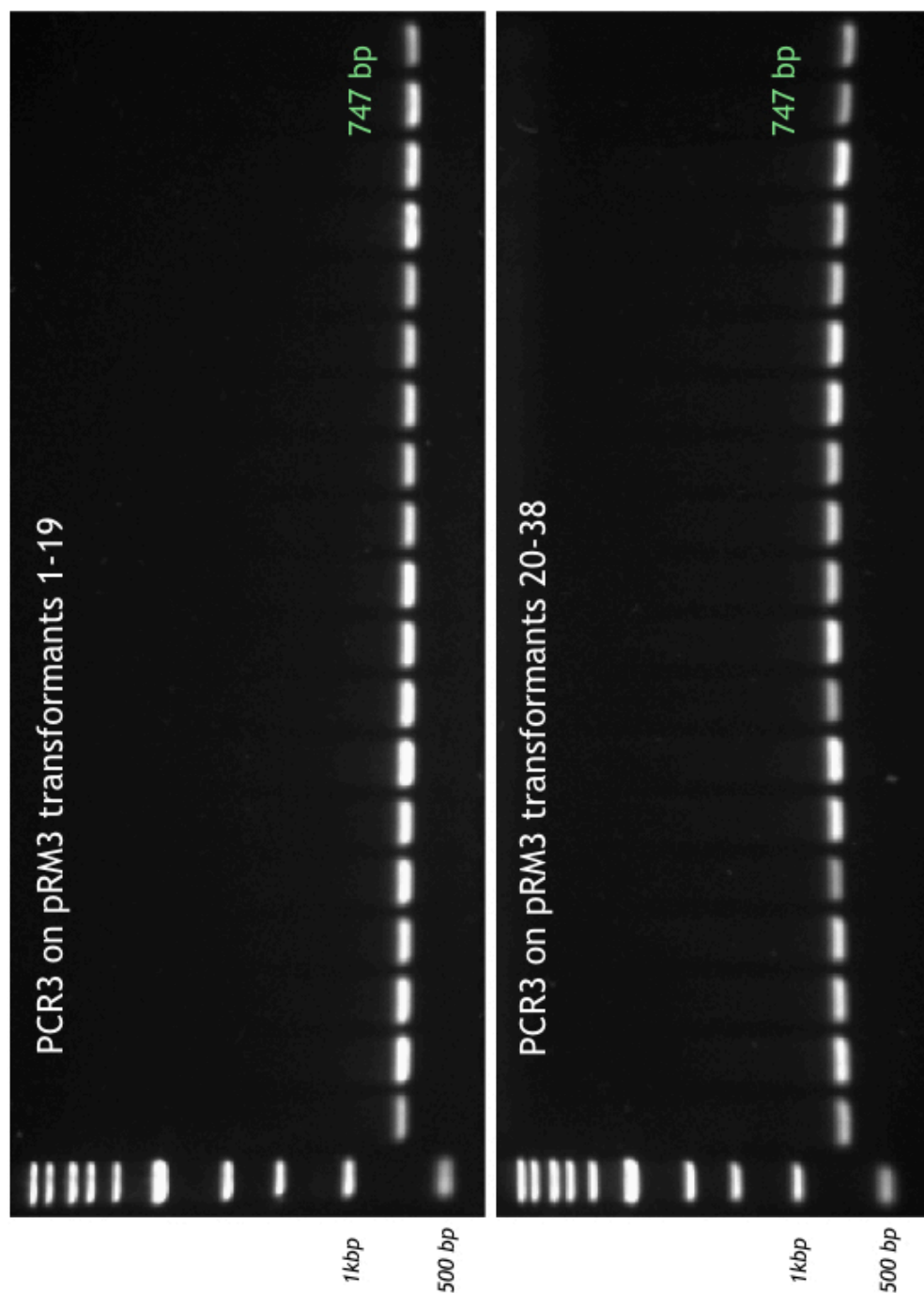


Figure 3.15: Agarose gels of 38 out of 76 confirmatory PCRs (Figure 3.14) conducted on colonies from LPC311a / pZJ7 + pRM3 transformation plates (LB + chloramphenicol). 76 out of 76 transformants were positive for this PCR across the first *att* site junction of the LP, suggesting that cassette exchange had been successful in 100% of pRM3 transformants.

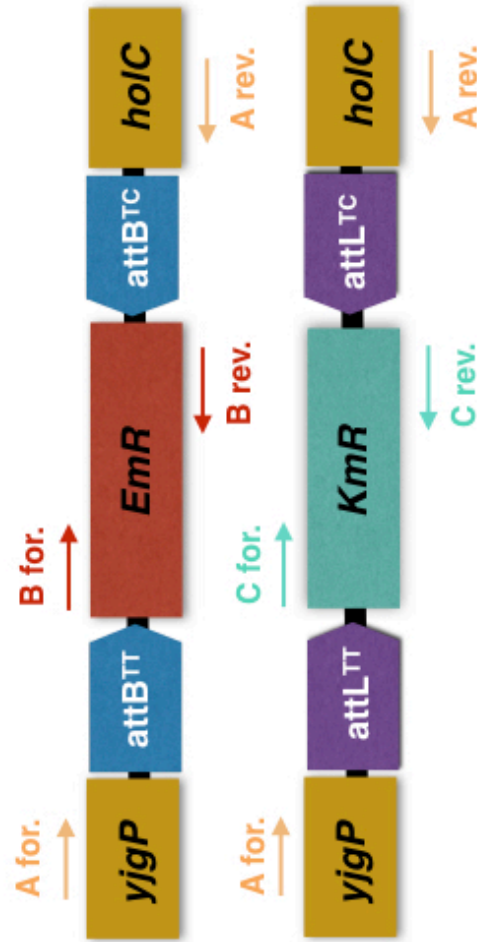
The PCR shown in Figure 3.14, which amplifies a sequence from upstream of *pepA* to inside the *CmR* marker, was used initially to test transformants in large numbers. Figure 3.15 shows an agarose gel of 38 of the 76 ara-induced LPC311a / pZJ7 + pRM2 transformants checked by this PCR. 100% of these transformants (76/76) showed a clear band of amplified DNA at the expected position (747 bp), indicating that *CmR* had been integrated into the LP at a high efficiency.

Further cassette exchange transformations were carried out with the *KmR* donor plasmid pRM2 in order to allow checks of antibiotic resistance marker gain and loss as described in section 3.5.2.1. The reason for this change from the above pRM3 transformations and PCR check was because selection for pZJ7 uses a chloramphenicol *CmR* marker, which would interfere with selection for the *CmR* donor cassette from pRM3. On the other hand, pRM2 features a *KmR* donor cassette (and *ApR* backbone), which does not clash with pZJ7.

In repeated transformations of LPC311a / pZJ7 cells with the *KmR* donor plasmid pRM2, transformants showed sustained resistance to all three marker antibiotics (erythromycin, chloramphenicol and kanamycin) following several rounds of heat treatment. To check if any cassette exchange or integration was occurring in these transformants, several PCRs were designed to read across all four of the possible junctions expected for either cassette exchange or no cassette exchange, and were performed on a smaller number of colonies. A diagram explaining these PCRs is shown in Figure 3.16.

According to expected band sizes indicated in the legend of Figure 3.16, successful cassette exchange should result in amplification from PCRs 3 and 4 and no amplification from PCRs 1 and 2, and *vice versa* for no cassette exchange. These PCRs were performed on six kanamycin-resistant LPC311a / pZJ7 pRM3-transformed colonies (arabinose-induced protocol), labelled A to F, and products were run on an agarose gel, shown in Figure 3.17.

Untransformed LPC311a was included as a cassette exchange negative control.



PCR	PRIMERS	PRODUCT	EXPECTED	OBSERVED (FIG 3.Q)
1. <i>pepA</i> flank to <i>EmR</i>	A for. & B rev.	1001 bp	✗	✗
2. <i>EmR</i> to <i>pepA</i> flank	B for. & A rev.	1331 bp	✗	✓
3. <i>pepA</i> flank to <i>KmR</i>	A for. & C rev.	924 bp	✓	✓
4. <i>KmR</i> to <i>pepA</i> flank	C for. & A rev.	1174 bp	✓	✗

Figure 3.16: Diagram representing four PCRs used in cassette exchange confirmations following transformations of arabinose-induced LPC311a / pZJ7 cells with pRM2. PCRs 1 and 2 are designed to read across the original ϕ C31 LP flanks, from upstream and downstream of *pepA* into the erythromycin resistance marker gene, and are expected to amplify in the case of no integration of the donor cassette. PCRs 3 and 4 are designed to read across the same flanks but into the *KmR* marker gene, and should amplify in the case of successful cassette exchange. Expected product sizes for each PCR are shown in the right-hand table, alongside an indication of the desired outcomes to confirm cassette exchange. The 5th column of this table corresponds to the observed products shown in Figure 3.17.

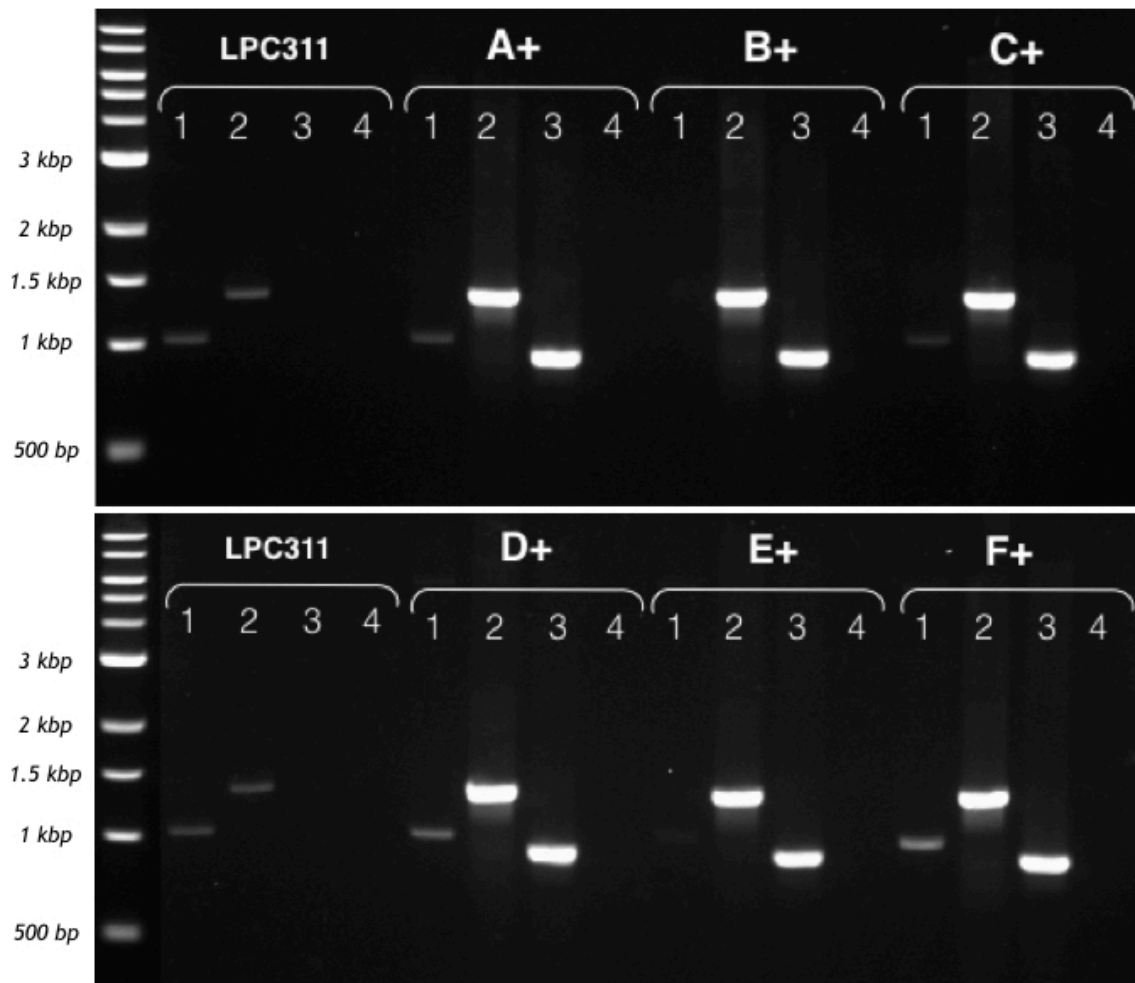


Figure 3.17: Agarose gel of PCRs as shown in Figure 3.16, performed on LPC311a (negative control) and 6 arabinose-induced LPC311 / pZJ7 + pRM3 transformants, labelled A+ to F+. The LPC311a negative control shows the expected bands for PCR1 and PCR2 of 1001 bp and 1331 bp, respectively, and no product for PCR3 and PCR4. Each of the six pRM3 transformants tested showed strong bands from PCRs 2 and 3, of sizes 1331 bp and 924 bp, respectively. This is different to what was expected from either successful or unsuccessful cassette exchange outcomes (as indicated in Figure 3.16). Complete integration of the *KmR* donor cassette should result in amplification in PCRs 3 and 4. From these results, it is clear that an alternative event to complete cassette exchange is occurring, resulting in the amplification of the upstream-*KmR* junction and the *EmR*-downstream junction of the landing pad.

Contrary to what was expected for the two expected possible outcomes, each pRM3 transformant amplified strongly for PCRs 2 and 3, with band sizes of 1331 bp and 924 bp, respectively. This indicated that the erythromycin resistance marker gene is still present in its original position at the *pepA* locus upstream of *attB^{TC}* and *holC*, while the kanamycin resistance marker has also been integrated at the start of the LP cassette.

An explanation which is consistent with these results is that a whole plasmid integration has occurred. In this outcome, pRM3 integrates via recombination between *attB^{TT}* on the genome and *attP^{TT}* on the donor plasmid, but the second expected recombination reaction between *attB^{TC}* and *attP^{TC}* does not occur. A diagram of this outcome is shown in Figure 3.18.

Whole integration of pRM3 into the *attB^{TT}* LP site not only explains the PCR results observed, but also explains why cells were persistently resistant to kanamycin, ampicillin and erythromycin following several rounds of heat treatment, as the plasmid and all three resistance markers were integrated into the genome and could not be removed as a temperature-sensitive plasmid would be.

Despite this undesired outcome being observed and complete cassette exchange being unsuccessful, the results shown in Figure 3.15 do reveal that recombination between *attB^{TT}* and *attP^{TT}* is highly efficient, and that integration of the whole plasmid at the *pepA* LP site has been observed at 100% efficiency in all transformants tested.

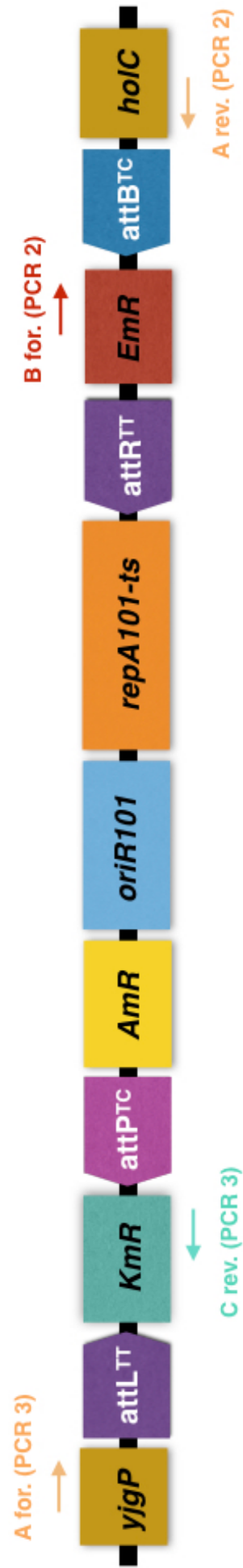


Figure 3.18: Diagram explaining observed results from Figure 3.17, showing complete integration of the pRM3 plasmid due to recombination between *attB^{TC}* LP and *attP^{TT}* donor cassette sites. This outcome is consistent with positive amplification of PCRs 2 and 3, indicated here by inclusion of primer locations for each.

3.5.3. Sequencing the ϕ C31 integrase LP on the LPC311a genome to investigate incomplete cassette exchange

Due to the absence of recombination between *attB^{TC}* on the genome LP and *attP^{TC}* on the donor cassette, the sequences of both were checked. Using primers reading upstream and downstream of *pepA* (PCR2 from Figure 3.4) to amplify the whole landing pad region, a PCR was performed, and the product was purified and sent for sequencing (Eurofins Genomics). pRM3 was also sent for sequencing to check the sequence of the donor cassette.

Sequences produced from the genomic LP PCR product were aligned with the desired sequence used to insert the LP into *pepA* by recombineering. This alignment is shown in Figure 3.19. From this alignment, it was revealed that a 17 bp deletion was present in *attB^{TC}*. This explains why recombination was not taking place between *attB^{TC}* and *attP^{TC}*, resulting in the integration of the whole pRM3 plasmid at attachment site *attB^{TT}*.

The next step of this study was to reestablish the ϕ C31 LP in the *pepA* locus and ensure that the complete sequence was present without any deletions or mutations. This would then allow efficient cassette exchange to be demonstrated, following on from the high rates of plasmid integration that have been seen on the LPC311a LP with only one complete *attB* site.

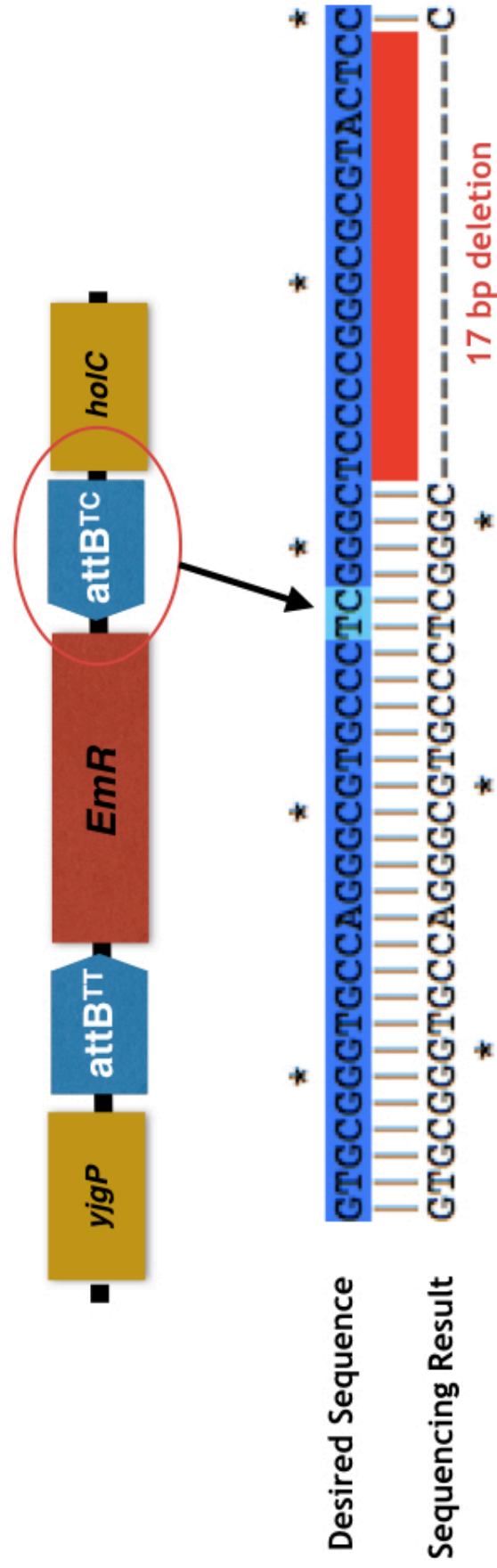


Figure 3.19: Results of sequencing from the *pepA* ϕ C31 LP site of LPC311a, aligned against the desired sequence of the LP. This shows that a 17 bp deletion is present in the second half site of *attB^{TC}*, explaining why the second round of recombination between *attB^{TC}* and *attP^{TC}* failed and resulted only in the complete integration of donor plasmids. Note that the *attB^{TC}* site is in an inverted orientation and therefore the top strand sequence shown is the reverse complement of the *attB* site described previously in this work. Therefore the 'TC' central dinucleotides in this diagram are shown as their reverse complementary pair 'GA', which are highlighted in red against the yellow of the *attB* site.

3.6. Discussion

In this project, a protocol was established and tested which allowed the delivery of a ϕ C31 integrase ‘donor cassette’ to a newly introduced site on the *E. coli* genome. This method made use of a temperature-sensitive delivery cassette which has not been used previously for serine integrase integration deliveries (Section 3.3; Methods Section 2.5).

A key aspect of the protocol was the expression of ϕ C31 integrase under control of the arabinose-inducible pBAD promoter. This overcame issues observed with constitutive expression of ϕ C31 integrase in landing pad cells which resulted in low cell growth and conversion of the ϕ C31 integrase expression plasmid to smaller plasmid products (Figures 3.9, 3.10 & 3.11). It is thought that this occurrence is connected to the presence of the *attB* sites on the genomic landing pad and a negative interaction with ϕ C31 integrase, as this modification of the expression plasmid was not observed in non-LP containing *E. coli* DS941 cells. However, once expression was controlled via the pBAD inducible promoter, recombination efficiency was much higher as demonstrated using the *Galk* recombination assay in LPC311a cells (Figures 3.12 & 3.13).

Initially, it was found difficult to successfully establish the LP cassette in place in the *E. coli pepA* gene due to very low efficiencies seen from the recombineering method. This low efficiency integration rate is a good illustration of why an improved method for chromosomal integration of heterologous DNA is necessary. Once marker-positive transformant colonies were seen on selection plates, there were only around three to five colonies per plate. It was later seen that this recombineering integration also introduced a loss-of-function deletion in the *attB^{TC}* attachment site of the LP (Figure 3.19), which also demonstrated the possible inaccuracies that can occur during recombineering. Unfortunately, this deletion within the LP attachment site was not noticed until late on in the project, meaning that a lot of time was spent trying to optimise a cassette exchange protocol despite the fact that the LP sequence was not as expected.

Taking into consideration the above limitation of the established LP only featuring one functioning *attB^{TT}* attachment site, the results of donor plasmid transformations (under arabinose-induced ϕ C31 integrase expression) can be viewed as whole plasmid integrations into a single chromosomally-integrated *attB^{TT}* site. Under this assumption, it can be argued that integration reactions were in fact highly efficient. Of every marker-positive transformant, PCR amplifications of the LP site revealed that whole plasmid integration had occurred, indicating that this event is highly efficient and demonstrable. Given that this event occurred at such a high rate, it is likely that, if a full secondary *attB^{TC}* (without deletion) were present, the second recombination reaction between *attP^{TC}* and *attB^{TC}* would occur (see Figure 3.18) equally efficiently to remove the remaining plasmid backbone and leave only the integration cassette remaining at the ϕ C31 LP site.

The next stage of this project was to demonstrate the described LP cassette exchange protocol with a *pepA*-located ϕ C31 integrase LP that contains the two desired complete *attB^{TT}* and *attB^{TC}* sites without any deletions. At the end of this project, recombineering was being repeated in order to establish the correct LP sequence. Further plans were made to create an LP for Bxb1 integrase in the same way. The design of this Bxb1 integrase LP was identical to that of the ϕ C31 integrase LP but with the ϕ C31 *attB^{TT}* and *attB^{TC}* sites replaced with 54 bp *attB^{GT}* and *attB^{GA}* sites for Bxb1 integrase, respectively, which are shown in Figure 3.20.

54 bp Bxb1 *attB* sites

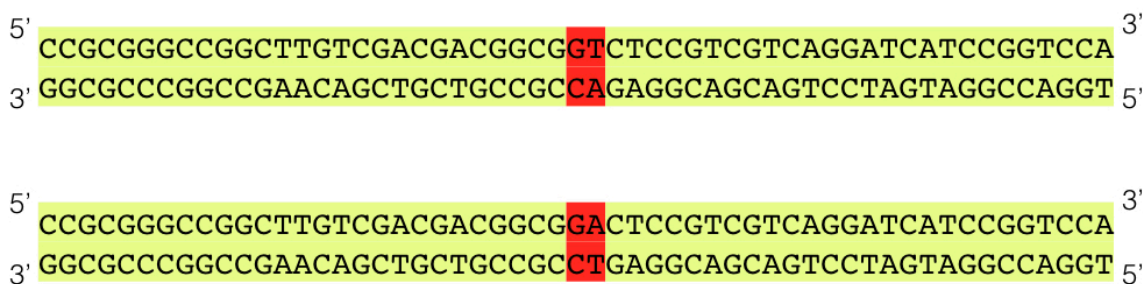


Figure 3.20: Bxb1 integrase *attB* sites to be used in Bxb1 LP construction similar to the design shown in Figure 3.1 but with Bxb1 attachment sites *attB^{GT}* and *attB^{GA}*.

Once LPs have been successfully demonstrated for cassette exchange further to the whole plasmid integrations shown here, these results can help to inform the design of a multi-integrase system in *E. coli*. Work done in the following chapters also aims to provide data on how LP cassette exchange efficiencies may differ between genomic locations (Chapter 4), whether there may be cross-reactivity problems when using multiple integrases (Chapter 5) and if this system can be successfully demonstrated in eukaryotic yeast cells (Chapter 6).

4. Chapter 4: Random transposition of several ϕ C31 LPs in *E. coli*

4.1 Introduction

So far in this project, a single gene locus, *pepA*, was targeted in order to establish and characterise LPs for ϕ C31 integrase and Bxb1 integrase in the same context. However, previous studies have shown that chromosomal location can have a significant effect on integration efficiencies. For instance, it has been observed that genomic integrations using a single TG1 integrase site are more efficient when placed close to the *E. coli* genome origin of replication, *oriC*, or the *E. coli* 'centromere' analogue site *migS*, than other areas of the genome (Muroi *et al.*, 2013). In order to determine the effect of chromosomal positioning on the efficiency of the ϕ C31 integrase LP, a library of *E. coli* strains with the LP cassette inserted into different places on the genome was created.

Expression levels in different genomic contexts are a factor in the design of reusable and tailored LP sites. An advantage of a multi-LP system is that expression levels are known and can be selected in order to fine-tune heterologous pathways. Engineering several LPs with different known natural levels of expression is also a convenient way to optimise heterologous gene or pathway expression without the need for switching promoter strength. Creating a large library of LPs with well characterised expression profiles would allow for an LP to be chosen based on its expression properties, an attractive prospect for integrating gene products that may need to be carefully regulated, for example in the case of toxic gene products whose levels need to be maintained below a certain threshold.

In order to determine the effect of chromosomal positioning on the characteristics of a ϕ C31 integrase LP, a library of *E. coli* strains was created via random transposition of the LP cassette across the entire *E. coli* genome. A ϕ C31 integrase LP library that was previously established in this lab (E. Conte, unpublished) was used as a basis for analysis in this project, which aimed to measure and compare cassette exchange efficiencies and heterologous gene expression levels between several random LP locations on the *E. coli* genome.

4.2 Design of ϕ C31 LP ISY100 transposition cassette and generation of *E. coli* LP library

The ϕ C31 integrase LP to be used in this experiment was designed to include both a selective antibiotic-resistance marker, *KmR*, and a counter-selective marker, *rpsL*, which confers sensitivity to streptomycin (Nair *et al*, 1993). These selective genes were flanked by *attB^{TT}* and *attB^{TC}* sites in a head-to-head orientation identical to the initial ϕ C31 integrase LP (*EmR*) previously designed in this project. The design of this LP is shown in Figure 4.1.



Figure 4.1: Diagram showing design of the ϕ C31 transposable LP featuring: a *KmR* marker for kanamycin resistance; *rpsL* counter-selectable marker for streptomycin sensitivity; flanking *attB^{TT}* and *attB^{TC}* landing pad sites for cassette exchange of an *attP^{TT}* and *attP^{TC}*-flanked target gene.

The transposase ISY100 was previously used in this lab to insert the *KmR* ϕ C31 LP into random positions in the *E. coli* DS941 genome (E. Conte, unpublished). The LP cassette was delivered via a plasmid (pZJ73; see Methods Section 2.5) into which it was cloned between two ISY100 terminal inverted repeat sequences. The ISY100 transposase expressed on the same plasmid facilitates integration of the transposable element, as defined by the ISY100 inverted repeat sequences, into TA dinucleotide sites on the genome, duplicating the TA upon transposition (Urasaki *et al*, 2002). The protocol for this plasmid transformation and library generation by random transposition can be found in Methods Section 2.18.

4.3 Determining the genomic locations of six randomly transposed ϕ C31 LPs by arbitrary PCR

From a library of around 1000 *KmR* ϕ C31 LP-positive *E. coli* (DS941) mutants, individual mutants were isolated by streaking out from library glycerol stocks onto LB agar plates containing kanamycin to obtain single colonies. Six mutants were selected for analysis and comparison of distant LP sites. In order to contextualise and characterise each of these landing pads within the *E. coli* genome, a two-step “arbitrary PCR” method was used to identify the location of each LP within the genome (see Methods Section 2.19). This arbitrary PCR is designed to amplify from a specific sequence on the inserted cassette to arbitrary sequences on the genome. An explanatory diagram of this arbitrary PCR is shown in Figure 4.2.

The first round of PCR used one primer that reads forward from a short sequence in the *KmR* gene of the LP cassette, and a second primer which consists of a sequence designed to anneal to a large number of possible sequences on the genome and a specific extension sequence (or tag). These primer sequences are shown in Methods Section 2.4 and the full method is described in Methods Section 2.19. This first arbitrary PCR produces a large number of products of varying sizes which have been amplified from different arbitrary locations on the genome to the LP *KmR* sequence. Using these products, a second PCR is performed using a forward primer which reads from just downstream of the first *KmR* primer, and a reverse primer which reads only from the extended (tag) sequence attached to the PCR products by the first arbitrary primer. The result is a cleaner and more concentrated PCR product that can be then sent for sequencing.

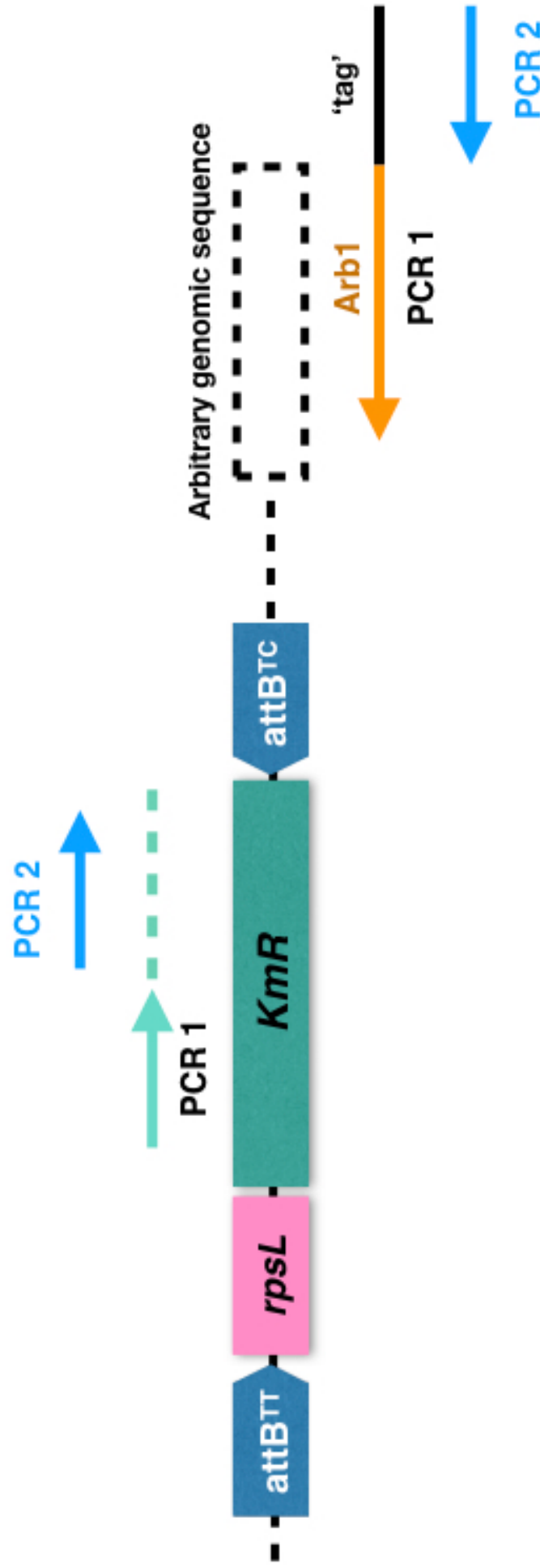


Figure 4.2: Diagram explaining the arbitrary PCR used to determine genomic locations of randomly transposed ϕ C31 LPs. The first round of PCR uses primers (labelled here PCR 1) to amplify from within the *KmR* gene (turquoise) to an arbitrary sequence on the *E. coli* genome, labelled Arb1 (orange), producing many lengths of PCR product. Arbitrary primer sequences are detailed in Methods Section 2.4. The products from this first PCR were then used as a template for a second PCR with primers (labelled PCR 2, blue) that were designed to read from within the amplified *KmR* gene to an extended sequence 'tag' created by the Arb1 primer from PCR 1. The products from this second PCR were then be sent for sequencing to determine the sequence identity of the genomic region downstream of the LP.

The amplified sequences from the arbitrary PCR were then sent for sequencing (Eurofins) and the results used to search the NCBI BLAST database for homology to a specific sequence on the *E. coli* substrain MG1655 (K-12), of which DS941 is a derivative. Results of the sequence alignments for the six *KmR* ϕ C31 LP-positive *E. coli* DS941 mutants tested are represented in Figure 4.3. The relative positions of each randomly transposed ϕ C31 LP on the *E. coli* genome is summarised in Figure 4.4, and the genomic context of each is detailed further in Table 4.1.

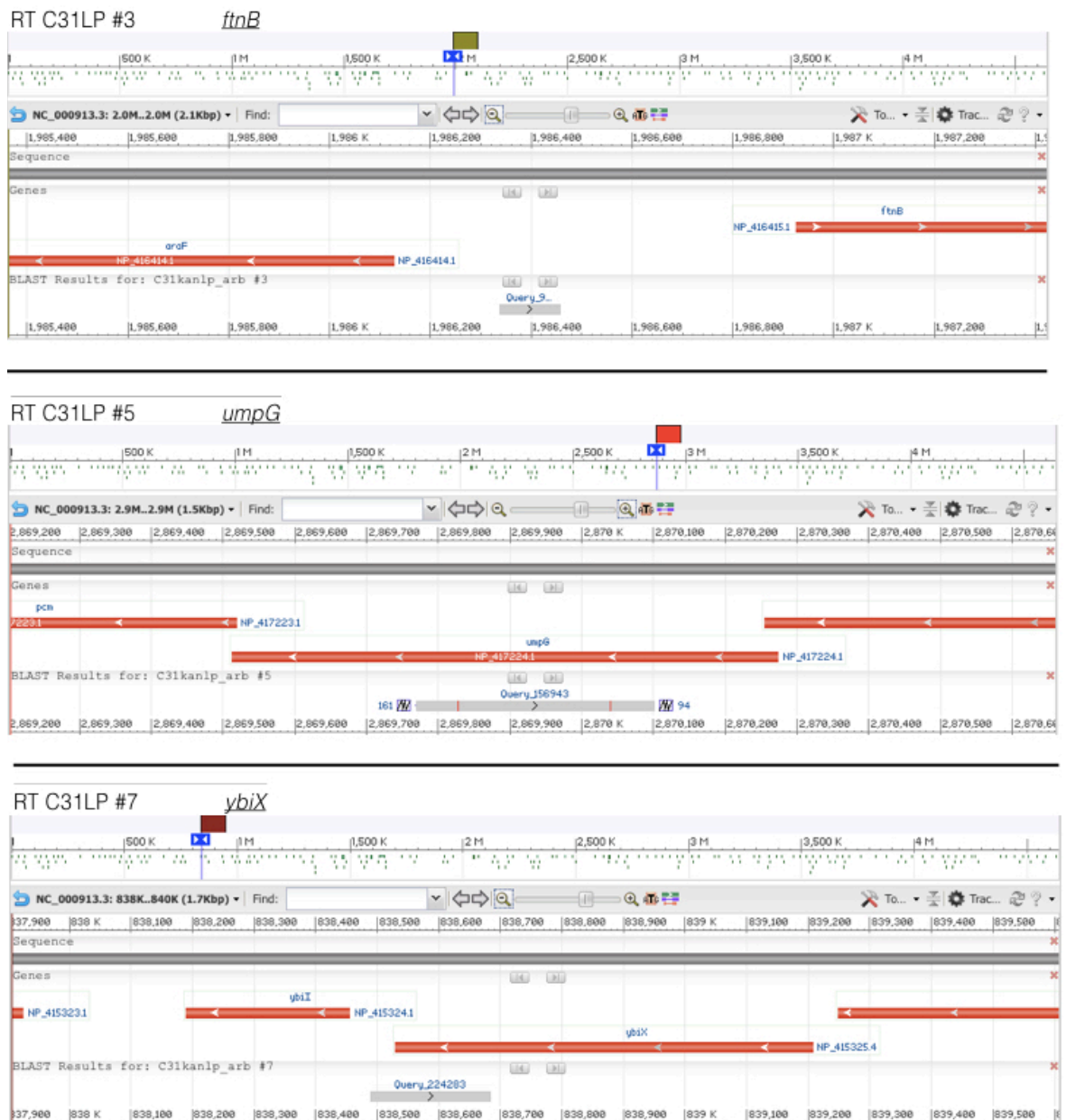


Figure 4.3a: BLAST sequence locations of three successful RTLP two-step arbitrary PCR product sequencing results, produced as alignments with the *Escherichia coli* str. K-12 substr. MG1655 reference genome in the NCBI genome explorer (https://www.ncbi.nlm.nih.gov/nucore/NC_000913.3). The top bar of each window indicates location of the sequence alignment within the 4,641,652 bp reference genome, while below this the PCR-derived sequence (which begins immediately at the end of the transposed LP sequence) alignment is shown in context of its region on the genome including genes in this region (scale of sequences in this diagram are indicated at the bottom of each window). The three mutants are labelled as follows according to their sample reference and the gene in closest proximity on the genome: RTC31LP3 *ftnB*; RTC31LP5 *umpG*; RTC31LP7 *ybiX*. References are shortened to ‘RTLP(n)’ elsewhere.



Figure 4.3b: BLAST sequence locations of three successful RTLP two-step arbitrary PCR product sequencing results, produced as alignments with the *Escherichia coli* str. K-12 substr. MG1655 reference genome in the NCBI genome explorer (https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3). The top bar of each window indicates location of the sequence alignment within the 4,641,652 bp reference genome, while below this the PCR-derived sequence (which begins immediately at the end of the transposed LP sequence) alignment is shown in context of its region on the genome including genes in this region (scale of sequences in this diagram are indicated at the bottom of each window). The three mutants are labelled as follows according to their sample reference and the gene in closest proximity on the genome: RTC31LP8 *ybdF*; RTC31LP9 *rutA*; RTC31LP10 *yhiM*. References are shortened to 'RTLP(n)' elsewhere.



Figure 4.4: Diagram indicating the relative positions of each RT φC31 LPs on the *E. coli* genome, named according to the gene in closest proximity on the genome. Starting position in reference to the NCBI *Escherichia coli* str. K-12 substr. MG1655 reference genome is labelled for each (see also Table 3.A). All LPs are in a left to right (5' to 3') orientation, as indicated by the position of each *attB^{TT}* (TT) and *attB^{TC}* (TC) site. The *E. coli* chromosomal origin of replication, *oriC*, is also labelled here for reference along with its position on the genome. Note: LPs lengths are not to scale with the *E. coli* genome.

Table 4.1: List of confirmed RT ϕ C31 LP mutants and their genomic locations, with reference to *E. coli* MG1655 (NCBI genome database).

Randomly Transposed ϕ C31 LP Mutant (closest gene)	Genome Position (BLAST <i>E. coli</i> MG1655)	Details of Gene (if inserted in coding region)
RTLP3 (<i>ftnB</i> ^o)	1986338	Non-coding region; upstream of <i>ftnB</i>
RTLP5 (<i>umpG</i> ^x)	2869769	<i>umpG</i> : broad spec. 5' (3')-nucleotidase and polyphosphatase; phenotype unknown
RTLP7 (<i>ybiX</i> ^o)	838490	Non-coding region; upstream of <i>ybdI</i>
RTLP8 (<i>ybdF</i> ^x)	605750	<i>ybdF</i> : DUF419 family protein; function unknown
RTLP9 (<i>rutA</i> ^x)	1073244	<i>rutA</i> : pyrimidine oxygenase; function in pyrimidine nitrogen catabolism; non-essential; loses ability to use pyrimidine as sole N source
RTLP10 (<i>yhiM</i> ^o)	3634499	Non-coding region; upstream of <i>yhiM</i>

4.4 Confirmation of genomic LP locations by location-specific PCR

In order to double-check that the genomic LP positions determined by arbitrary PCRs and sequencing were correct, PCRs were designed with primers reading from the *KmR* marker on the ϕ C31 LP to a 20 bp site on the genome under 1000 bp from the assumed LP location. The primers used are listed in Methods Section 2.4 and the PCR was done using an NEB Phusion Polymerase protocol (Methods Section 2.13.1) on a single colony of each RT LP mutant (diluted in distilled water before addition to reaction tube). DNA produced was cleaned using a QIAgen PCR Purification kit before being run in an agarose gel. This agarose gel is shown in Figure 4.5.

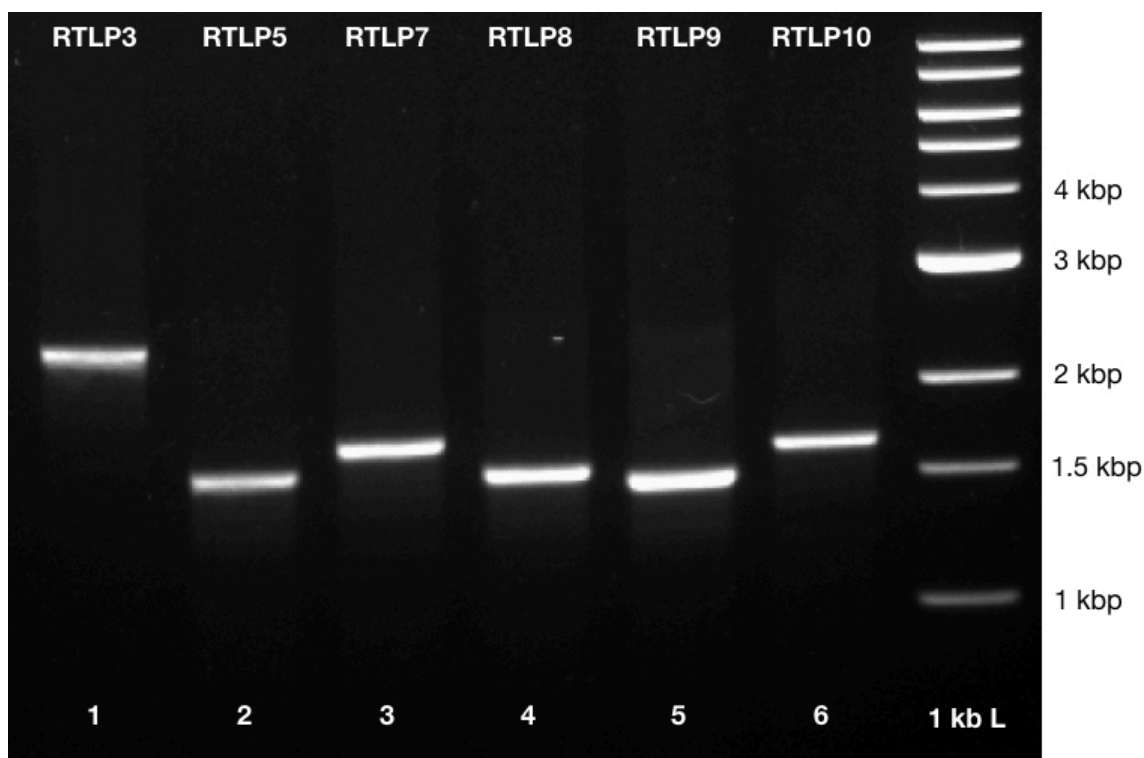


Figure 4.5: Agarose gel showing the DNA produced in confirmation PCRs of randomly transposed ϕ C31 LPs (*KmR*) for each of the six mutants tested. The expected product sizes for each LP mutant were as follows: RTLP3 (*ftnB*) 2313 bp; RTLP5 (*umpG*) 1568 bp; RTLP7 (*ybiX*) 1860 bp; RTLP8 (*ybdF*) 1670 bp; RTLP9 (*rutA*) 1507 bp; RTLP10 (*yhiM*) 1700 bp.

4.5 Designing and cloning a temperature-sensitive donor cassette plasmid compatible with *KmR* LP

Cassette exchange experiments conducted with the randomly transposed *KmR* ϕ C31 LPs required the cloning of a new temperature-sensitive donor cassette. In this case, the erythromycin resistance gene *EmR*, used as an LP marker in previous experiments, was cloned into the ϕ C31 LP donor cassette of plasmid pRM1 (*AmR* backbone) between restriction sites *NheI* and *NotI* to create a new temperature-sensitive donor plasmid, pRM4 (detailed in Methods Section 2.5). Cassette exchange transformations can now be done as previously using the inducible ϕ C31 integrase expression plasmid pZJ7 as described in Chapter 3.

4.6 Discussion

Although cassette exchange experiments were not carried out during this work, the project has established and confirmed the genomic locations of six randomly inserted ϕ C31 integrase LPs from the wider library previously created in this lab (E. Conte, manuscript in preparation). LP positions were first identified using a genome-wide arbitrary PCR and then confirmed using specific PCRs with primers that read forward from a site upstream of the identified LP site and reverse from a sequence within the *KmR* gene. This contextual information can inform cassette exchange experiments to determine whether location on the genome has an effect on cassette exchange efficiency at the LP site.

The six LPs identified represent a range of contexts. Three are integrated within the coding regions and three are in non-coding regions, and they are spread across the entire chromosome. This last consideration is important as it will be interesting to see whether LP cassette exchange efficiency will be increased nearer the chromosomal origin of replication, *oriC*, as this has been observed previously in studies that integrated plasmid DNA into single sites using TG1 integrase (Muroi *et al.*, 2013).

Cassette exchange experiments would use the same protocol developed in the previous chapter, expressing ϕ C31 integrase under control of an arabinose-induced promoter and transforming a temperature-sensitive donor vector, pRM4, described above which features an *EmR* marker flanked by *attP^{IT}* and *attP^{TC}* sites. Cassette exchange would then be detected by first selecting for erythromycin resistance, treating cells at 42 °C to remove the donor plasmid, then checking to see the antibiotic resistances of the resulting cells in large numbers to determine what percentage of transformant colonies were successful for ϕ C31 integrase-mediated cassette exchange. In these experiments, the *rpsL* counter-selectable marker can also be used to check if cassette exchange has taken place. If cells are sensitive to streptomycin, unsuccessful cassette exchange is indicated due to persisting presence of *rpsL* on the LP. However, if cells are streptomycin-resistant, this indicates that the

rpsL has successfully been replaced by the donor cassette on the LP. Finally, several transformants would be tested for cassette exchange using PCRs, as done in Chapter 3.

After cassette exchange reactions with donor marker genes have been done to determine the comparative efficiency at each site, the effect of LP position on the expression levels of an integrated gene can be explored. The ability to choose from a number of genomic LP sites with different expression profiles would be of advantage to the proposed LP library system, as it would allow the user to select an LP to suit the desired expression levels of the target gene. Additionally, in a multi-integrase system, being able to choose LPs with higher or lower expression levels would be of interest in pathway optimisation when integrating several pathway genes.

To test expression levels of each LP, a donor cassette consisting of a green fluorescent protein (GFP) gene flanked by *attP^{TT}* and *attP^{TC}* sites for ϕ C31 integrase would be introduced. If cassette levels of the LPs prove to be suitably high, there would be no need for a selective marker on this donor cassette. It will then be possible to dilute and plate GFP-transformed cells (without marker selection) to achieve single colonies. A high number of these colonies should then produce GFP, which will be detectable using a fluorescence imager. Using an imager which is able to quantifiably measure fluorescence levels will give an initial indication of expression level. However, measuring single colonies with a single copy of GFP on the chromosome will not accurately represent expression levels. A single GFP-positive colony could be used to inoculate an overnight starter culture of which the cell density would be measured, and then a known number of cells could be used to inoculate a fresh cell culture. This culture could then be used to measure fluorescence levels of a known number of cells in liquid culture, and levels could be compared between LP locations.

5. Chapter 5: Investigating compatibility of integrases for a multi-LP system

5.1 Introduction

One of the most promising applications of the landing pad system is the use of several integrase LPs to allow multiplexed gene integrations into a single strain. Given the range of integrases now available for use, it may be possible to create a mutant with many serine integrase LPs (e.g. ϕ C31.int, Bxb1.int, ϕ BT1.int, TG1.int), into which several target genes can be integrated to introduce an entire metabolic pathway. Further possibilities are presented by the use of recombination directionality factors (RDFs), or integrase-RDF fusions, to perform additional rounds of cassette exchange. This chop-and-change method could for example be used to analyse the effect of expressing an alternative or optimised gene within an already-integrated pathway, and would limit the steps required for strain optimisation.

Serine integrases are by nature highly site-specific, yet there have been suggestions of cross-reactivity, of attachment site recognition and RDF interactions, between different integrases given the close relatedness of these enzymes (Singh *et al.*, 2014). For example, ϕ C31 integrase and TG1 integrase share around 49% sequence identity (Rowley *et al.*, 2008), which has led to suggestions that this level of shared sequence could result in cross-interaction between these two integrases and their non-corresponding RDFs. For a multi-integrase LP system to be successful, cross-reactivity of the integrases and their RDFs must be investigated in order to address any cross-compatibility issues. As well as providing results to inform the design of a multi-integrase LP system, the knowledge of to what extent serine integrases interact with each other's attachment sites and RDFs is of use to the development of other systems, in particular for the serine integrase recombinational assembly (SIRA) (Colloms *et al.*, 2013).

Two of this group's most-studied serine integrases, ϕ C31 integrase and TG1 integrase, are here investigated for the cross-compatibility of the integrase enzymes and their corresponding RDF proteins. TG1 integrase is a serine

integrase that has been shown to efficiently integrate heterologous DNA into pre-inserted *att* sites on the *E. coli* genome (Muroi *et al.*, 2013), making it a prime candidate for use in the landing pad system. As previously mentioned, ϕ C31 and TG1 integrase also share a high level of primary sequence identity, meaning that they are two enzymes from which some cross-interaction might be expected. The aim of this study was to determine to what extent ϕ C31 integrase and TG1 integrase and their RDFs, (gp3 and gp25 respectively), interact, and whether they could be used effectively within the same multi-integrase system. Following this, the extent of attachment site orthogonality will be measured to determine whether it is possible for these serine integrases to recombine closely related *att* sites which feature high sequence identity.

5.2 Measuring cross-compatibility of ϕ C31 and TG1 integrases and their corresponding RDFs

5.2.1. ϕ C31 and TG1 integrase *attL* x *attR* recombination activation by non-native RDFs

Serine integrases act to recombine *attP* and *attB* sites to integrate DNA and form product sites *attL* and *attR* in a highly unidirectional manner; this means that when only the integrase is present, no recombination occurs between *attL* and *attR*, and the reaction favours only integration. The reverse reaction (*attL* recombination with *attR*) proceeds only in the added presence of the integrase's RDF. It is thought that RDF proteins bind to integrase subunits to end inhibition of the *attL/attR* reaction, and this integrase-RDF complex catalyses *attL/attR* recombination. This reverse recombination reaction usually requires the combination of an integrase and its own RDF. However, it has been suggested that some integrases may be activated by RDFs of other integrases, due to the closely related nature and presence of conserved sequences in these RDFs.

Here, the integrases ϕ C31 and TG1 and their respective RDFs gp3 and gp25 are investigated for such cross-interactions in *attL/attR* recombination. This involves incubating each integrase and its native or non-native RDF with an

attL/attR substrate to determine to what extent opposite RDFs are able to facilitate this reaction compared to native RDFs. These reactions were first tested *in vitro* using a recombination assay similar to the one used in the integrase activity assay in Chapter 3 (see Section 3.4.1). Substrate plasmids in this case feature *attL* and *attR* sites and are designed for *in vitro* reactions; pFM52, featuring *attL* and *attR* sites for ϕ C31 integrase and pFM139, featuring *attL* and *attR* sites for TG1 integrase. These plasmids are detailed in Methods Section 2.5. A diagram describing this *in vitro* recombination assay is shown in Figure 5.1.

For each integrase, four reactions were set up: 1) no integrase, replaced by buffer IDB as negative blank; 2) integrase only; 3) integrase plus own RDF; 4) integrase plus non-native RDF. Recombination assay reactions were first set up by pre-incubating the integrase and RDF for 15 minutes at room temperature before adding the plasmid substrate (a full protocol is described in Methods Section 2.20). Initial reactions were run for either 30 minutes or 3 hours to determine whether incubation time had an effect on recombination efficiency. However, when results indicated that there was no discernible difference in result with a longer incubation, 30 minute incubation periods were used for all consequent reactions. Once reactions were completed, *NruI* was added to create linear products and DNA was run on an agarose gel.

From photos of these agarose gels, relative DNA concentrations of the products of each reaction were quantified using ImageJ (see Methods Section 2.23). This method gives the relative percentage of the DNA concentration of each band in each lane. Unrecombined DNA is seen as two bands at the top and bottom of each lane, while recombination results in two bands between these unrecombined bands (see Figures 5.1, 5.2 & 5.3 for more detail).

Figures 5.2 and 5.3 show the agarose gels of the above *in vitro* recombination assays carried out with ϕ C31 integrase and gp3 or gp25 (Figure 5.2), and TG1 integrase with gp25 or gp3 (Figure 5.3), alongside graphs illustrating the recombination levels quantified from each gel lane according to Methods Section 2.23.

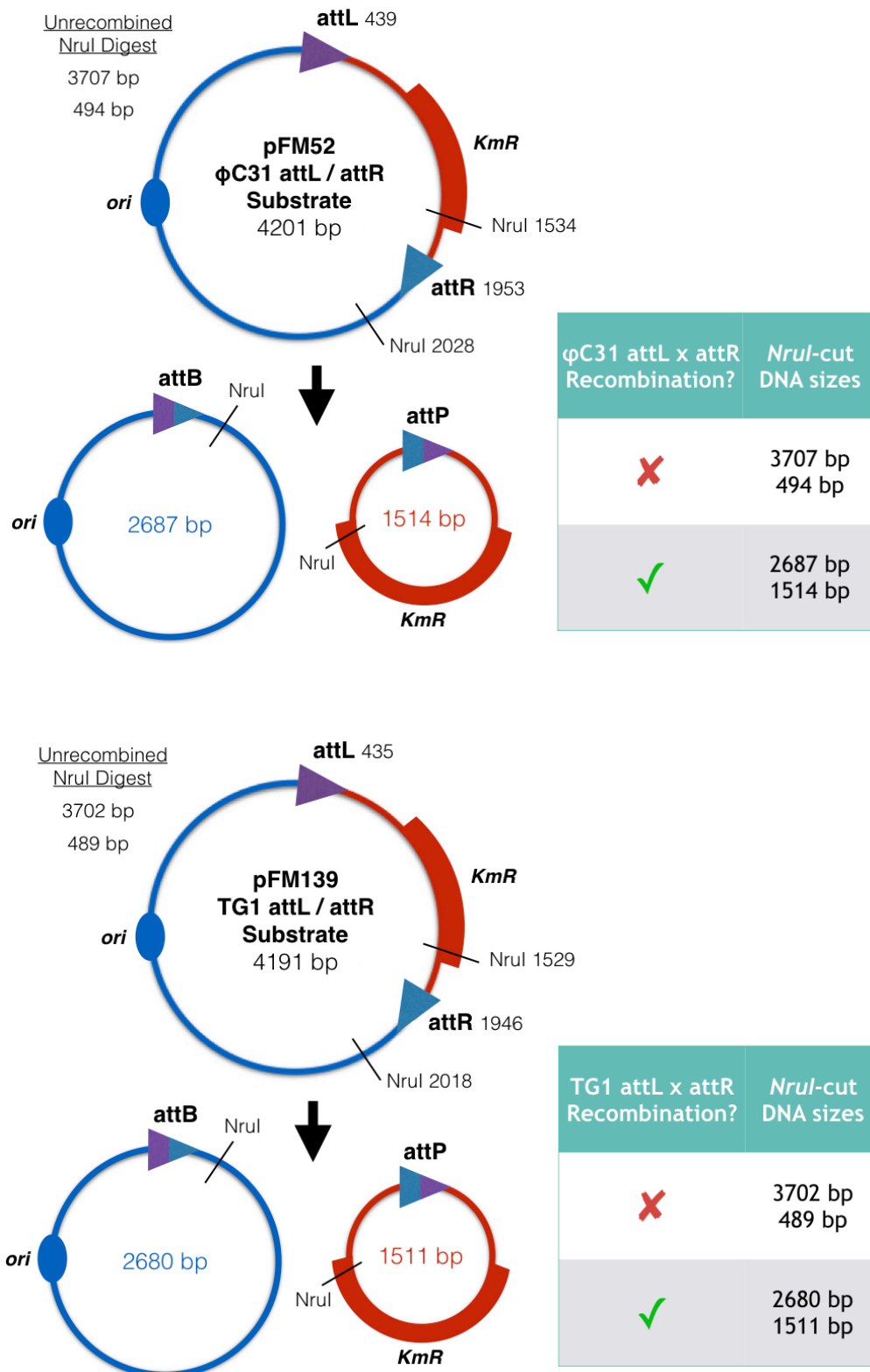


Figure 5.1: *In vitro* assay for *attL* x *attR* recombination by φC31 integrase (pFM52, top) and TG1 integrase (pFM139, bottom). When incubated with purified integrase and RDF *in vitro*, *attL* x *attR* recombination results in two circular products featuring an *attP* or *attB* site, which are both linearised by the restriction enzyme *Nrul*. In the case that no recombination between *attL* and *attR* occurs, the original substrate plasmid is cut to form two linear products of the sizes indicated.

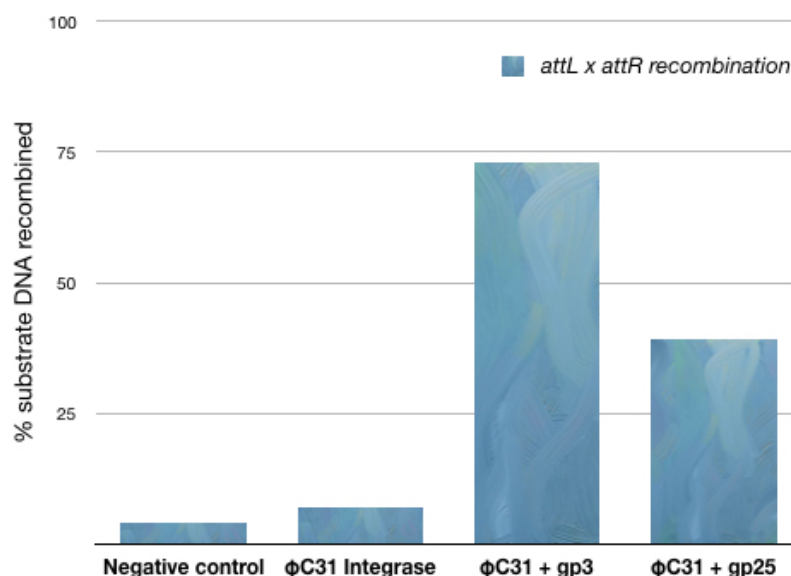
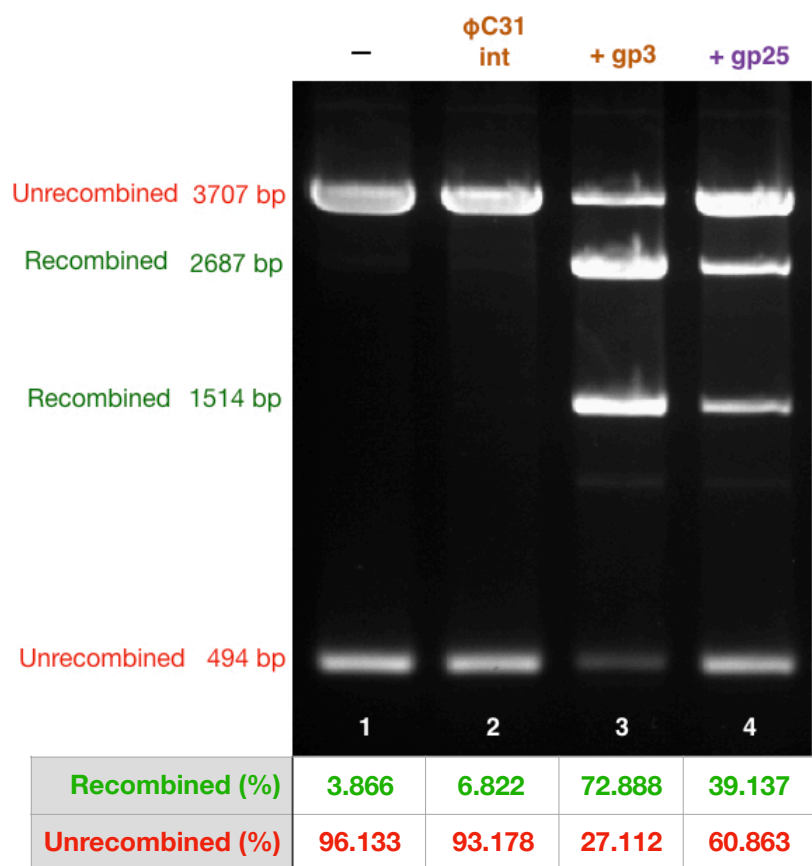


Figure 5.2: Agarose gel showing DNA from the *in vitro* *attL* x *attR* recombination assay of ϕC31 integrase with either native RDF gp3 or TG1 RDF gp25. The four lanes of the gel are as follows: 1) substrate pFM52 with IDB buffer as a negative control; 2) pFM52 with ϕC31 integrase only; 3) pFM52 with ϕC31 integrase incubated with gp3; 4) pFM52 with ϕC31 integrase incubated with gp25. Figures below the gel image correspond to the percentage of substrate DNA in the gel that was recombined or unrecombined, calculated with ImageJ (see Methods 2.23). The percentage of substrate DNA recombined in each reaction is represented graphically below. These figures are estimates produced using computation analysis of the agarose gel image and do not represent precise DNA concentration measurements.

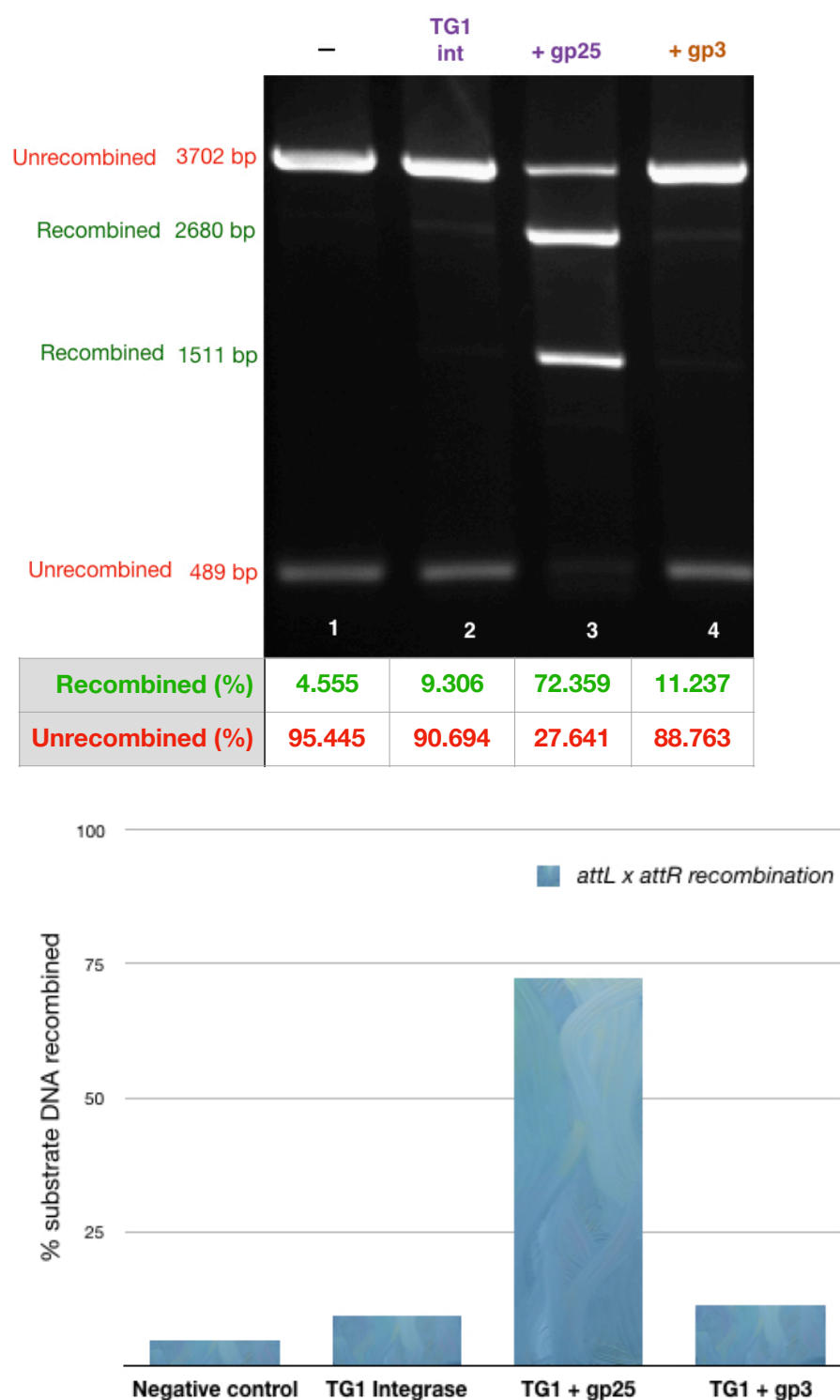


Figure 5.3: Agarose gel showing DNA from the *in vitro* *attL* x *attR* recombination assay of TG1 integrase with either native RDF gp25 or ϕ C31 RDF gp3. The four lanes of the gel are as follows: 1) substrate pFM139 with IDB buffer as a negative control; 2) pFM139 with TG1 integrase only; 3) pFM139 with TG1 integrase incubated with gp25; 4) pFM139 with TG1 integrase incubated with gp3. Figures below the gel image correspond to the percentage of substrate DNA in the gel that was recombined or unrecombined, calculated with ImageJ (see Methods 2.23). The percentage of substrate DNA recombined in each reaction is represented graphically below. These figures are estimates produced using computation analysis of the agarose gel image and do not represent precise DNA concentration measurements.

It would naturally be expected that only an integrase's native RDF would be able to activate *attL* x *attR* recombination. However, results observed here show that TG1 gp25 is able to significantly activate ϕ C31 integrase *attL* x *attR* recombination. The agarose gels of Figure 5.2 show that incubating ϕ C31 integrase with gp25 (Figure 5.2, lane 4) produces a large amount of recombination product, with 39% of the substrate DNA recombined according to digital quantification of the DNA bands. However, it should be noted that this recombination is less efficient than the natural gp3 activation of ϕ C31 (Figure 5.2, lane 3), which recombined 73% of the substrate DNA.

Interestingly, the reverse does not appear to be true, as ϕ C31 gp3 did not activate TG1 integrase *attL* x *attR* recombination at the same level (Figure 5.3, lane 4), with only 11% of the substrate plasmid recombined. This number is very close to the recombination level seen in the reaction containing only TG1 integrase without an RDF (9% recombination; Figure 5.3 lane 2). This demonstrates the low level of *attL* x *attR* recombination that TG1 integrase is capable of even in the absence of an RDF.

5.2.2. Inhibition of ϕ C31 and TG1 integrase *attP* x *attB* recombination by ϕ C31 gp3 and TG1 gp25 RDFs

In addition to activating *attL* x *attR* recombination, RDFs have been shown to have an inhibitory effect on the reverse *attP* x *attB* reaction (Khaleel *et al.*, 2011). After finding that TG1 gp25 is able to activate ϕ C31 integrase *attL* x *attR* recombination, the next step was to determine to what extent these RDFs exerted an inhibitory effect on *attP* x *attB* recombination for each integrase. This was again done using an *in vitro* recombination assay almost identical to that above but instead using substrate plasmids featuring *attP* and *attB* sites. The plasmids in this case were pFM16 (ϕ C31 *attP/attB*) and pFM138 (TG1 *attP/attB*), which are detailed further in Methods Section 2.5. A diagram detailing this *in vitro* recombination assay and product sizes is shown in Figure 5.4.

The *in vitro* recombination assay was carried out as before, with a 15 minute pre-incubation of the integrase and RDF at room temperature before adding the substrate plasmid and incubating for 30 minutes (detailed further in Methods Section 2.20). After the reaction was stopped, NruI was added and the resulting linear products were run on an agarose gel, which are shown in Figures 5.5 (ϕ C31 and substrate pFM16) and 5.6 (TG1 and substrate pFM138), alongside graphs illustrating the recombination levels quantified from each gel lane according to Methods Section 2.23.

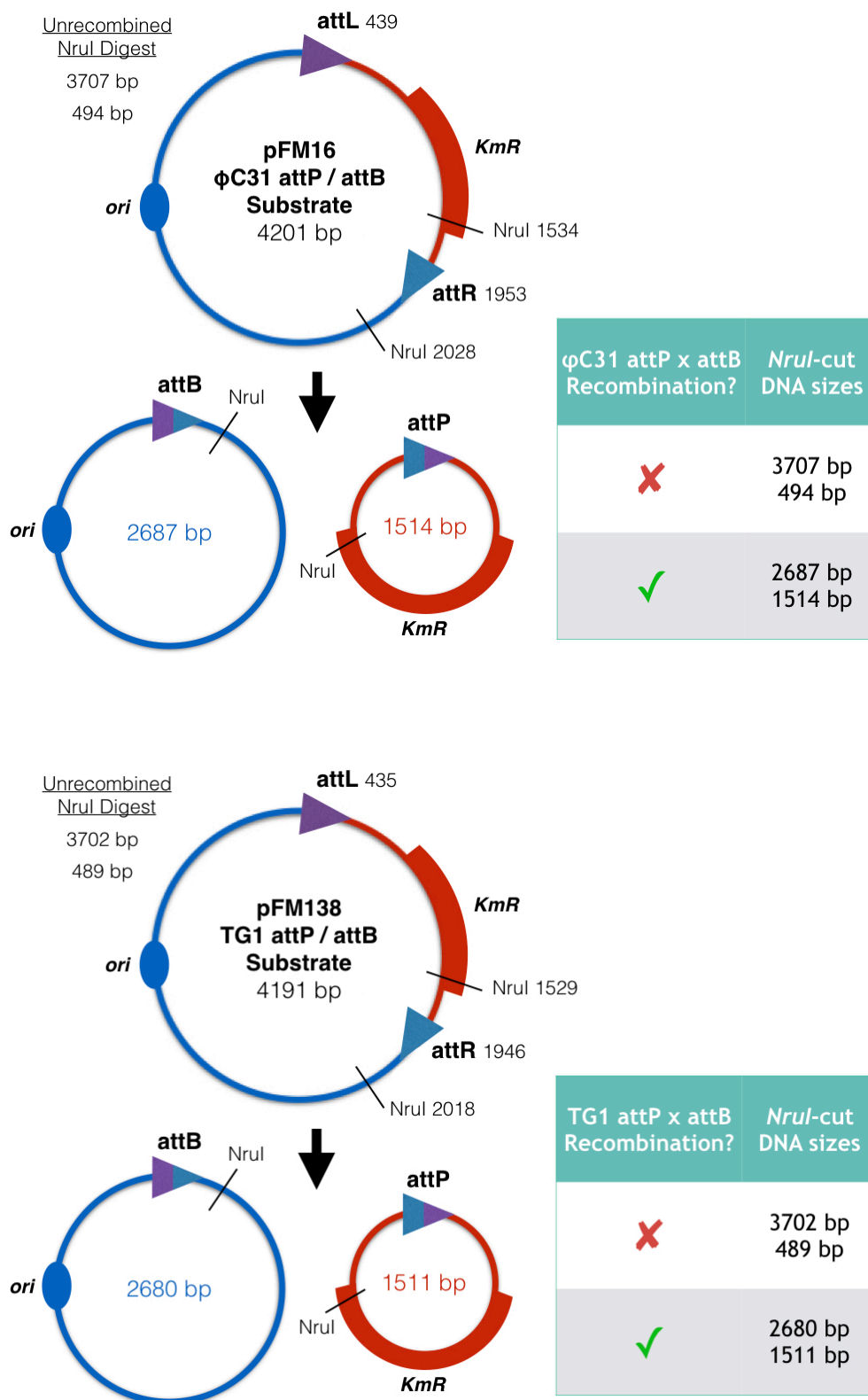


Figure 5.4: *In vitro* assay for *attP* x *attB* recombination by φC31 integrase (pFM16, top) and TG1 integrase (pFM138, bottom). When incubated with purified integrase *in vitro*, *attP* x *attB* recombination results in two circular products featuring an *attL* or *attR* site, which are both linearised by the restriction enzyme Nrul. In the case that no recombination between *attP* and *attB* occurs, the original substrate plasmid is cut to form two linear products of the sizes indicated.

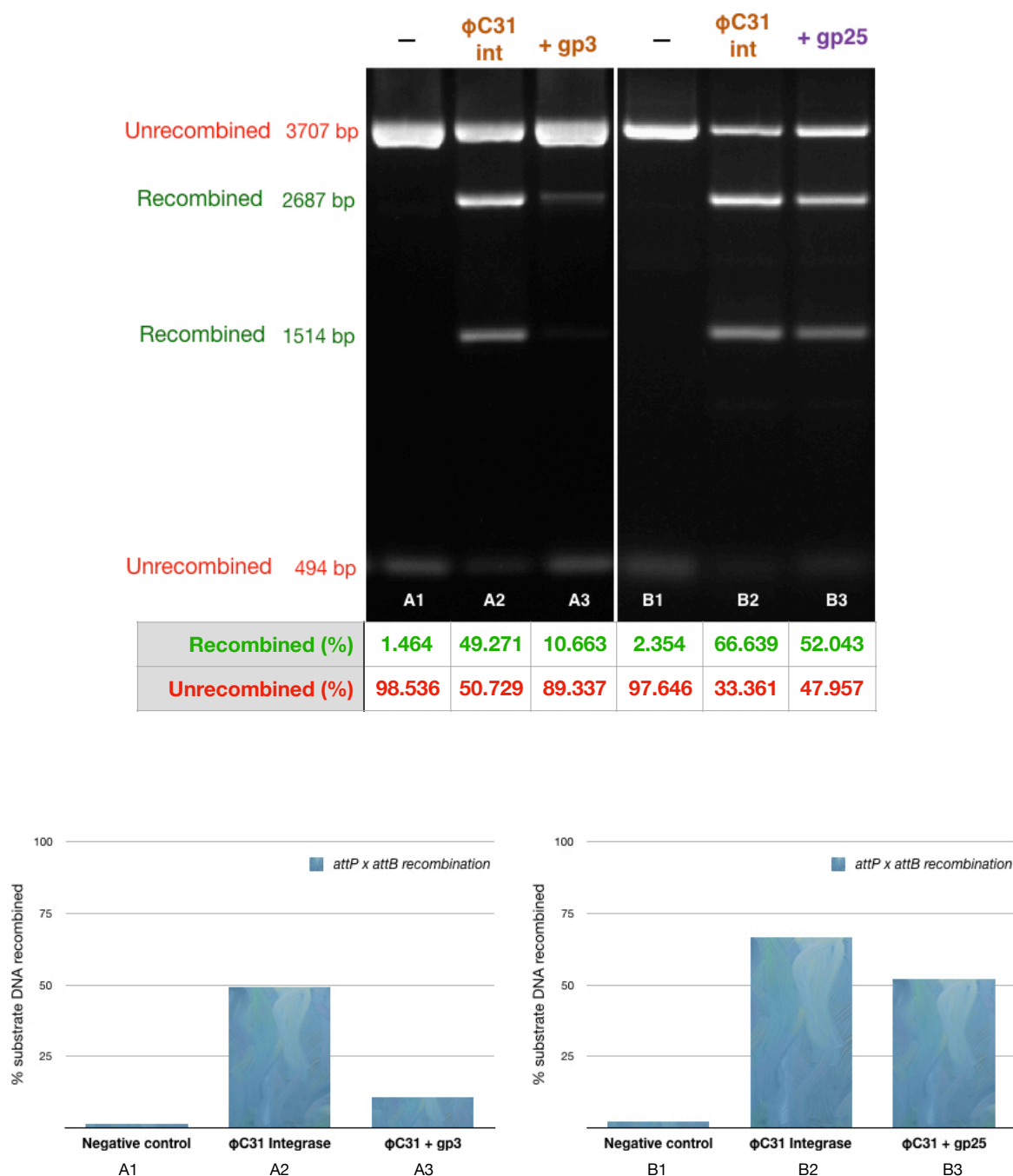


Figure 5.5: Agarose gel showing DNA from the *in vitro* *attP* x *attB* recombination assay of ϕ C31 integrase with either native RDF gp3 or TG1 RDF gp25. Two gels are shown, A and B, with lanes as follows: A1) substrate pFM16 with IDB buffer as a negative control; A2) pFM16 with ϕ C31 integrase only; A3) pFM16 with ϕ C31 integrase incubated with gp3; B1) substrate pFM16 with IDB buffer as a negative control; B2) pFM16 with ϕ C31 integrase only; B3) pFM16 with ϕ C31 integrase incubated with gp25. Figures below the gel image correspond to the percentage of substrate DNA in the gel that was recombined or unrecombined, calculated with ImageJ (see Methods 2.23). The percentage of substrate DNA recombined in each reaction is represented graphically below. These figures are estimates produced using computation analysis of the agarose gel image and do not represent precise DNA concentration measurements.

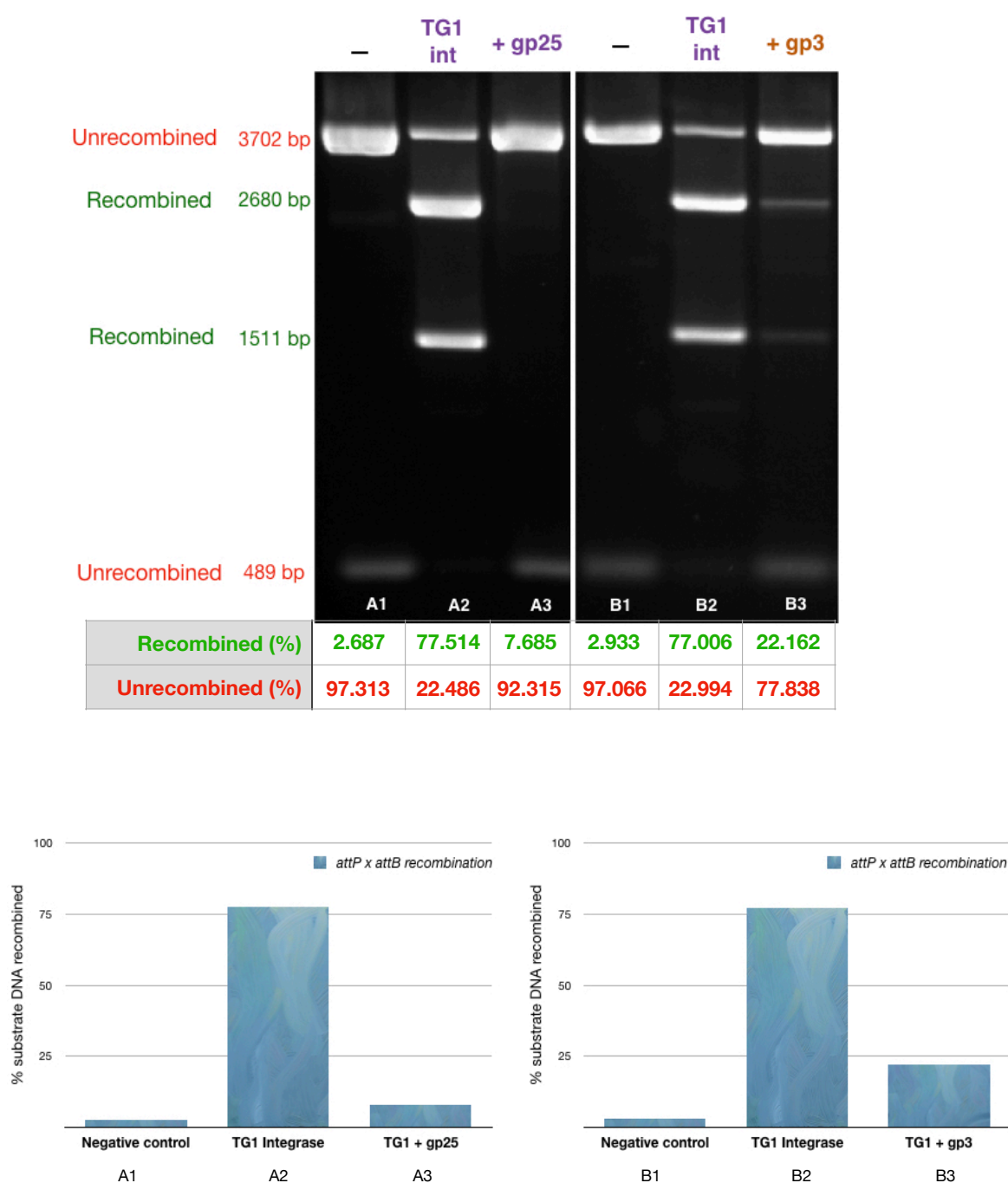


Figure 5.6: Agarose gel showing DNA from the *in vitro attP x attB* recombination assay of TG1 integrase with either native RDF gp25 or ϕ C31 RDF gp3. Two gels are shown, A and B, with lanes as follows: A1) substrate pFM138 with IDB buffer as a negative control; A2) pFM138 with TG1 integrase only; A3) pFM138 with TG1 integrase incubated with gp25; B1) substrate pFM138 with IDB buffer as a negative control; B2) pFM138 with TG1 integrase only; B3) pFM138 with TG1 integrase incubated with gp3. Figures below the gel image correspond to the percentage of substrate DNA in the gel that was recombined or unrecombined, calculated with ImageJ (see Methods 2.23). The percentage of substrate DNA recombined in each reaction is represented graphically below. These figures are estimates produced using computation analysis of the agarose gel image and do not represent precise DNA concentration measurements.

The *attP* x *attB* recombination assay results shown in Figures 5.5 and 5.6 firstly demonstrate how each integrase's own RDF inhibits *attP* x *attB* recombination. This activity is particularly evident in TG1 integrase. When its RDF gp25 is added, recombination of the *attP* and *attB* substrate was shown to decrease from 76% to 7% (Figure 5.6, gel A), indicating a significantly strong inhibition of the reaction. While for ϕ C31 integrase this inhibition was not a strong, adding gp3 to the reaction still had an inhibitory effect, decreasing recombination from 49% to 11% (Figure 5.5, gel A).

Interestingly, while TG1 RDF gp25 was shown in Figure 5.2 to activate ϕ C31 integrase *attL-attR* recombination (and the reverse was not true, with ϕ C31 gp3 shown not to activate TG1 integrase *attL-attR* recombination; Figure 5.3), in the case of *attP-attB* inhibition, gp25 appeared not to inhibit the forward reaction significantly (Figure 5.5, lane B3). Unexpectedly, it was ϕ C31 gp3 that had a more significant effect on TG1 *attP-attB* recombination inhibition, with its addition reducing the level of substrate DNA recombined from 77% to 22% (Figure 5.6, lanes B2 & B3).

5.3 Discussion

The results presented here illustrate the high unidirectionality of ϕ C31 and TG1 integrase when recombining *attP* and *attB* sites. Without the presence of gp3 or gp25, integrase enzymes show no or very little recombination of integrase product sites *attL* and *attR* (Figures 5.2 & 5.3). The naturally high level of recombination between *attP* and *attB* for both ϕ C31 and TG1 integrase can be seen in Figures 5.5 and 5.6, for which integrase and substrate were only incubated for a total of 30 minutes before stopping the reaction. Additionally, the low activity of each integrase on *attL* x *attR* sites without RDF demonstrates how the integration reaction occurs in a strongly unidirectional manner. This characteristic of the serine integrases is important for their application in the landing pad system as it ensures that high efficiency of integration can be achieved without any reverse flow of the

reaction. High efficiency of cassette exchange will allow the landing pad-mediated genomic integration method to forgo the rate-limiting selection and screening process during strain development.

Concerning the reverse reaction between *attL* and *attR*, both integrases were shown to display a high level of recombination when pre-mixed with their own RDFs (Figures 5.2 & 5.3), as expected. However, when ϕ C31 integrase was mixed with the TG1 RDF gp25, it also displayed a significant level of *attL* x *attR* recombination (Figure 5.2). Previous amino acid sequence alignments between ϕ C31 and TG1 integrase have revealed that they share a sequence identity of 63.8% in the catalytic N-terminal of the enzyme (49.5% identity overall) (Morita *et al.*, 2009). It is apparent that interaction is occurring between ϕ C31 integrase and TG1 gp25, however this interaction is not similarly seen between TG1 integrase and ϕ C31 gp3, suggesting that each of these RDFs interact slightly differently with the integrase enzyme. However, further binding experiments will be required to determine the nature of these protein interactions. Current evidence suggests that the RDF interacts with the coiled-coil domain of the integrase C-terminal, a region which is thought to facilitate formation of the synaptic complex and control directionality of *attB/attP* and *attL/attR* recombination (Rutherford & Van Duyne, 2014).

Inhibition of *attP* x *attB* recombination by addition of RDF was seen with both integrases and their native RDFs. This effect was particularly strong in the case of TG1 integrase, with addition of gp25 completely preventing *attP* x *attB* recombination (Figure 5.6, lane A3). This result supports the potential of TG1 integrase and gp25 being used for secondary cassette switching in a multi integrase/RDF LP system, in which a second target gene could be switched in (following expression of the integrase and its RDF) using the *attL/attR* sites produced by initial cassette exchange. The product sites of this secondary cassette exchange could then be used to switch in another *attP* donor cassette, and so on. However, it will be important to consider the observation that gp25 is capable of facilitating ϕ C31 integrase *attL* x *attR* recombination in any multi-integrase system, as there could be complications of using ϕ C31 and TG1 integrase for *attL* x *attR* reactions in the same cells.

An interesting and unexpected result is the inhibition of TG1 integrase *attP-attB* recombination by ϕ C31 RDF gp3. It would be reasonable to expect that, due to gp25's activating activity of *attL-attR* recombination by ϕ C31 integrase (and the lack of activation of TG1 *attL-attR* recombination by gp3), it would be gp25 that would have the higher inhibitory effect on ϕ C31 integrase. However, what was observed is that ϕ C31 gp3 has a much higher inhibitory effect on TG1 integrase *attP-attB* integrase. However, it is worth noting that the relationship between RDF concentration on *attL-attR* activation and *attP-attB* inhibition was not explored extensively in this work. Further dosage effect experiments, in which a very low concentration of RDF is added and increased in small increments over time, may be necessary to determine the true inhibitory effect of each RDF on other integrases.

While these experiments show that ϕ C31 integrase interacts with TG1 gp25 *in vitro*, further experiments need to be done to determine to what extent this occurs within the cell. It will be important to determine whether ϕ C31 gp3 interacts with TG1 integrase *in vivo* despite this cross-reactivity not being apparent *in vitro*. The basis for this thinking is that this interaction may be more likely to take place *in vivo* as a larger concentration of RDF would be accrued when expressing the protein from a plasmid in the cell. This increased concentration of RDF might be capable of tipping the equilibrium to favour *attL* x *attR* recombination. If this were the case, control of both integrase and RDF expression by inducible promoters might be necessary.

Another candidate serine integrase for use in an LP system is ϕ BT1. However, there is also strong evidence that ϕ BT1 gp3 and ϕ C31 gp3 are interchangeable *in vitro* and *in vivo* (Zhang *et al.*, 2013), with both RDFs sharing an 85% sequence identity. Given the results from this study, it will be interesting to observe whether a similar cross-interaction occurs between ϕ BT1 integrase, TG1 integrase and their RDFs. It has also been suggested that ϕ BT1 integrase is capable of recombining *attL* and *attR* sites at a low level in the absence of ϕ BT1 gp3. This will need to be investigated if designing a ϕ BT1 LP, as it would reduce the high efficiency required for markerless cassette exchange.

A significant basis of the multiple landing pad system and other multi-integrase systems, such as serine integrase recombinational assembly (SIRA), is the specificity of an integrase for its own attachment sites. Due to similarities in their *att* site sequences, it will be important to establish the fidelity of each integrase to its own *att* sites and not those of other integrases. To this end, recombination experiments should be done both *in vivo* and *in vitro*, combining each integrase with *attP/attB* substrates from other integrases. Thus it will be seen to what extent an integrase is able to recombine *att* sites of another enzyme, and whether these integrases could be successfully used within the same multiplexed system.

Further experiments to characterise serine integrases for use in a multi-LP system should also focus on the orthogonality of crossover sites within *att* sites. The LP system relies on the paradigm that *att* sites recombine only with other *att* sites containing identical central dinucleotide crossovers. Mechanistically, non-complementary central pairings should not be able to ligate with one another, preventing the final ligation step of recombination. However, *in vivo*, due to the existence of DNA repair mechanisms, it may be possible for cells to ‘fix’ the mismatching problem and proceed with recombination. Therefore, for each integrase, *in vitro* and *in vivo* recombination assays should be carried out (Methods Sections 2.20 & 2.21) using plasmid substrates containing an *attP* and *attB* site with non-matching central dinucleotides.

6. Chapter 6: Demonstrating a landing pad in *Saccharomyces cerevisiae*

6.1. Introduction

While demonstrating and establishing that the serine integrase landing pad system in *Escherichia coli* has useful applications in strain development and optimisation for biotechnology, demonstrating the same system in eukaryotic cells is of equal or perhaps greater significance. Several industrially important microorganisms are eukaryotic and therefore present a more complex cellular context for genomic cassette integrations. The yeast *Saccharomyces cerevisiae* is a keystone of the biotechnology industry and has long been used as a host for producing commercial enzymes, biopharmaceuticals and biofuels, due to its well-understood characteristics and long history of use in the fermentation industry. This makes *S. cerevisiae* a key candidate for the further expansion and development of the landing pad system, as well as leading the way for the export of the LP to other eukaryotic industrially-relevant microorganisms such as *Pichia pastoris*.

In addition, demonstrating the use of serine integrases in eukaryotic cells could provide a basis for further research into the use of these enzymes in more complex multicellular eukaryotes for applications such as creation of transgenic lab animals or for gene therapy. ϕ C31 integrase has already been demonstrated to function successfully in several experimentally relevant organisms, with studies conducted in *Drosophila* (Groth *et al.*, 2004, Bischof *et al.*, 2007, Bateman *et al.*, 2006); however most of these studies relied on the existence of *att*-like psuedosites on the genome. ϕ C31 integrase has also similarly been used in mammalian cells such as mice and human embryonic stem cells (Thyagarajan, *et al.*, 2001), either using pseudo-sites or single inserted *attP* sites for integration of whole plasmids.

Employing the landing pad system in *S. cerevisiae* will both establish a highly useful tool for strain development and begin the development of a potential multi-LP yeast strain similar to that described for *E. coli* earlier in this thesis. Characterising the LP platform in this eukaryotic microorganism will also determine whether the same system could feasibly be used in yet more complex multicellular eukaryotes such as *Drosophila*, mice, or mammals used in experimental biology. In this chapter, a ϕ C31 integrase LP tailored for use in *S. cerevisiae* was designed and targeted to several locations on the yeast genome. These LPs will be spread across a single yeast chromosome in order to determine the effect of location, for example proximity to the centromere or telomeres, on integration efficiency.

6.2. Establishing a ϕ C31 LP at several loci on the *S. cerevisiae* genome

6.2.1. Designing a *S. cerevisiae* ϕ C31 LP

The ϕ C31 integrase landing pad to be used in *S. cerevisiae* consists of the same *attB^{TT}* and *attB^{TC}* sites in a head-to-head orientation as the *E. coli* LP; however, the marker chosen is changed for use in yeast. In this case, the marker gene is a yeast auxotrophic selection marker, *URA3*. The *URA3* marker gene can be used in yeast strains which are auxotrophic for uracil production and ordinarily require addition of uracil to the growth medium in order to grow. When *URA3* is introduced into a uracil-negative strain, its ability to produce uracil is restored and it is able to grow on media lacking uracil. The *URA3* marker is also chosen here for its counter-selectable property. Loss of *URA3* can be detected by growth on media containing 5-fluoroorotic acid (5-FOA). This is possible because the gene product of *URA3*, orotidine 5'-phosphate decarboxylase, converts 5-FOA to a toxic product, 5-fluorouracil, resulting in cell death. Therefore *URA3* was chosen as a versatile marker which can be used for detection of the LP when established, and later for counter-selection during LP-donor cassette exchange experiments. A diagram of the ϕ C31 integrase *S. cerevisiae* LP is shown in Figure 6.1.



Figure 6.1: Diagram representing the design of a ϕ C31 integrase LP for use in *S. cerevisiae*, featuring the *URA3* auxotrophic selection marker flanked by ϕ C31 *attB^{TT}* and *attB^{TC}* sites in a head-to-head orientation.

6.2.2. Selecting genomic locations for LP integration

This project makes use of a library of yeast knockout strains established by the *Saccharomyces cerevisiae* Genome Deletion Project (Cherry *et al*, 2012). The Genome Deletion Project was set up to facilitate phenotypic analysis of every open reading frame in the yeast genome, and yielded a library of knockout strains in which each open reading frame has been replaced by the *KanMX* (kanamycin-resistant) deletion cassette. This is of use to this LP project as it presents the opportunity to be able to insert the LP cassette into any gene in the yeast genome via homologous recombination using the same *KanMX*-targeting homology sequences. From the Genome Deletion Project library, five mutant strains were selected with the *KanMX* knockout located at sites spread evenly across Chromosome IV. A table describing these genes and their locations is shown in Table 6.1 and their estimated relative positions represented in Figure 6.2.

These gene locations were chosen on the basis that their knockouts are viable and have no or little competitive disadvantage. One gene was selected in the centromeric region (G9/YDR007W) in order to investigate whether proximity to the centromere affects integration efficiency. Four of these knockouts (G9/YDR007W, D2/YDR138W, E11/YDR297W, G5/YDR539W) are located in a region of Chromosome IV with a high concentration of transposable elements; it is thought that this might effect the efficiency of LP integrations due to this region being naturally more physically open and available to transposases, for example having longer stretches of nucleosome-free chromosomal DNA.

Table 6.1: *KanMX* gene knockouts chosen to be targeted for insertion of the ϕ C31 integrase LP cassette with details of mutant phenotypes (Cherry *et al*, 2012).

Working Reference	SGD Reference	Details of Gene & Mutant Phenotype	Chr IV Coordinates
F8	YDL229W	<i>SSB1</i> : Cytoplasmic ATPase, chaperone. HSP70 protein involved in heat-shock. $\Delta(SSB1,SSB2)$ = cold-sensitivity.	44065 to 45906
G9	YDR007W	<i>TRP1</i> : Phosphoribosylanthranilate isomerase. Part of tryptophan biosynthesis pathway. $\Delta TRP1$ = tryptophan auxotrophy & cold-sensitivity.	461842 to 462516
D2	YDR138W	<i>HPR1</i> : Subunit of THO/TREX complexes which maintain telomeres and regulate lifespan. $\Delta HPR1$ = decreased anaerobic growth rate.	730578 to 732836
E11	YDR297W	<i>SUR2</i> : Sphinganine C4-hydroxylase. Catalyses a step in sphingolipid biosynthesis. $\Delta SUR2$: No significant loss of function.	1056551 to 1057600
G5	YDR539W	<i>FDC1</i> : Ferulic acid decarboxylase. Converts carboxylic acids to vinyl derivatives. $\Delta FDC1$: No significant loss of function.	1512094 to 1513605

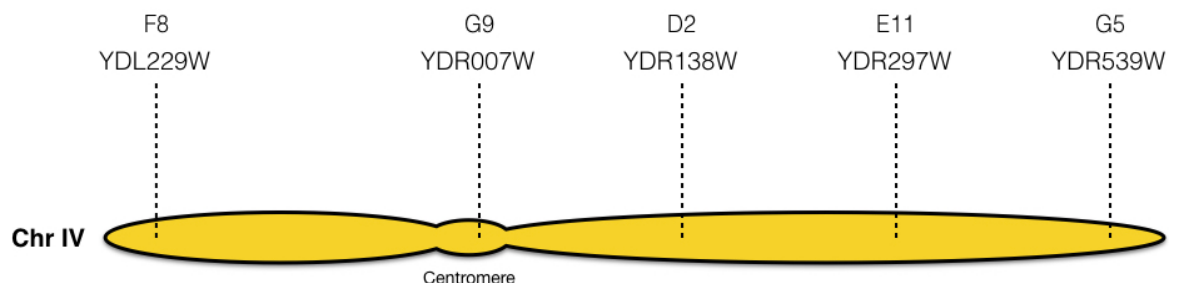


Figure 6.2: A diagram representing the approximate relative positions of each of the 5 genes selected on *S. cerevisiae* Chromosome IV.

6.2.3. Inserting the ϕ C31 LP into *KanMX* marker library strains by homologous recombination

6.2.3.1 Cloning of ϕ C31 LP *KanMX* knockout delivery plasmid

The method chosen for inserting the ϕ C31 integrase LP into the *KanMX* yeast mutants was homologous recombination. This makes use of the endogenous DNA repair mechanism and requires the introduction of the desired LP cassette with flanking homology sequences to the target *KanMX* gene. When transformed into the cell, homologous sequences are recombined, with the *KanMX* cassette being replaced by the LP cassette. This mechanism is represented in Figure 6.3.

The *URA3* ϕ C31 LP with flanking *KanMX* homology was introduced into yeast cells for homologous recombination via a delivery plasmid. The *E. coli* cloning vector pMTL23 was selected as a delivery plasmid, which features a ColE1 origin and ampicillin resistance gene. As this plasmid was only being used for delivery of the cassette and didn't need to be propagated in *S. cerevisiae*, it was sufficient for all cloning steps to be done in *E. coli*. This also overcame the problem of removing the delivery plasmid after transformation. The regions of homology to be used for homologous recombination were chosen as 120 bp at the very start (positions 1-120) and 120 bp at the very end (positions 1238-1358) of *KanMX*. These sequences were synthesised as a single gBlock oligonucleotide (from Integrated DNA Technologies), featuring restriction sites for cloning flanking and separating the two homology sequences, shown in Figure 6.4.

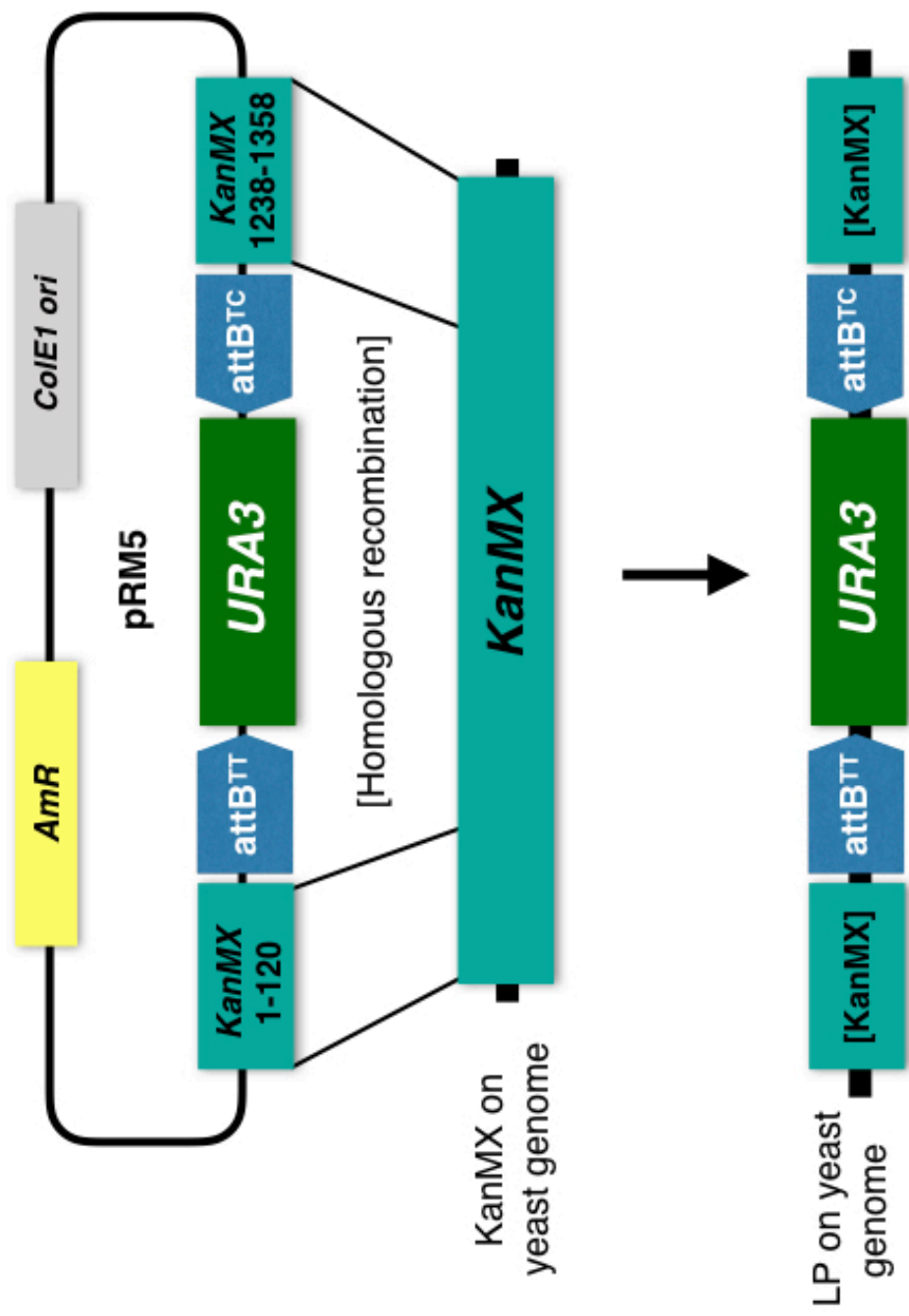
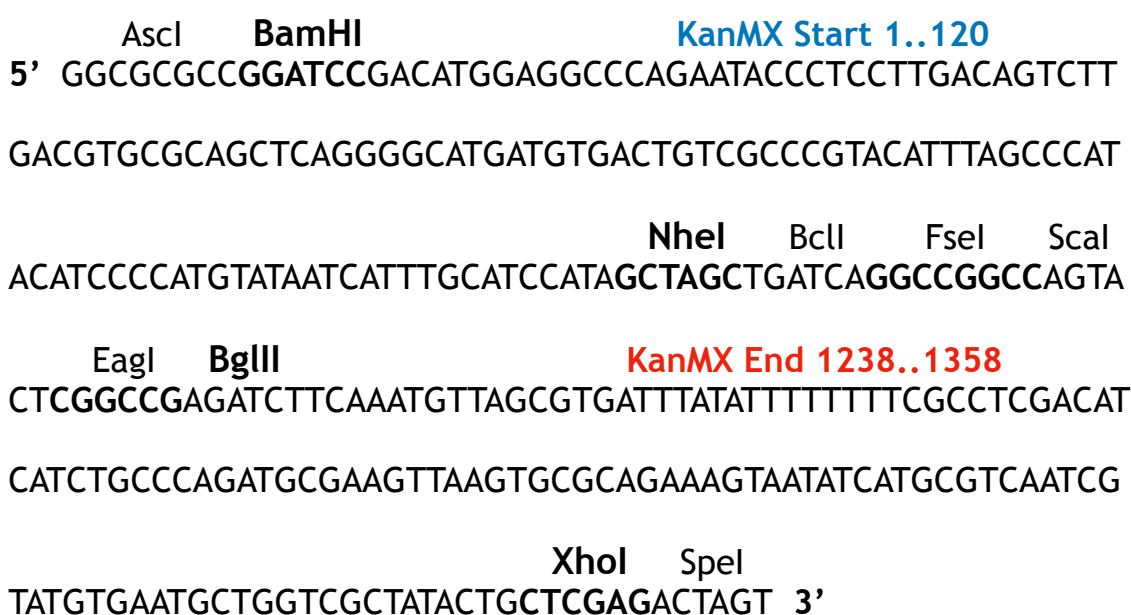


Figure 6.3: The ϕ C31 integrase LP replaces the *KanMX* cassette by homologous recombination according to homologous regions flanking the LP.



153

The *KanMX* homology gBlock was cloned into pMTL23, replacing the existing polylinker, using restriction enzymes BamHI and XhoI. Restriction mapping was used to confirm the insertion of the gBlock fragment, followed by sequencing using forward primer M13-for upstream of the replaced polylinker and reverse primer M13-rev reading 3' to 5' from the end of the cloned site. This intermediate plasmid is referred to as pMTL23-*KanMX* for the following steps.

The *URA3* ϕ C31 integrase LP was taken from a previously cloned LP cassette (see Methods Section 2.5), which features the ϕ C31 LP sequence as shown in Figure 6.1, flanked by restriction sites *NheI* upstream of *attB^{TT}* and *BglII* downstream of *attB^{TC}*. In order to use these restriction sites for correct cloning of the LP cassette into pMTL23, it was necessary to remove an additional *NheI* restriction site that was present in the pMTL23 backbone. This was done by partial digest using the restriction enzyme *BmtI*, an *NheI* isoschizomer which cuts to leave a 3' overhang, allowing for end filling with Pfu polymerase and blunt-ended re-ligation to remove the *NheI* site. The removal of the *NheI* site was confirmed by restriction mapping and consequent sequencing.

The yeast *URA3* ϕ C31 LP delivery plasmid, designated pRM5, was consequently constructed by digesting the landing pad cassette from pYLP1 and pMTL23-*KanMX* vector construct using restriction enzymes *NheI* and *BglII*. The digested LP insert and pMTL23-*KanMX* vector were then purified by agarose gel electrophoresis and ligated to produce pRM5, shown in Methods Section 2.5.

6.2.3.2 Replacing *KanMX* with a *URA3* ϕ C31 LP by homologous recombination

ϕ C31 LP *KanMX* replacement plasmid pRM5 was introduced into each of the five chosen *KanMX* yeast knockout strains via a lithium acetate yeast transformation method described in Methods Section 2.22. Transformed cells were plated onto SD-uracil (SD media minus uracil) plates to select for *URA3*. To test for replacement of the *KanMX* gene with the *URA3* ϕ C31 LP cassette via homologous recombination, transformants were subsequently patched onto two selective plates. Firstly, to further test for presence or absence of *URA3*, cells were patched onto SD-uracil agar where growth would only be possible under expression of *URA3*, therefore confirming presence of the *URA3* LP. To test whether the *KanMX* cassette had been knocked out from the genome and replaced by the ϕ C31 LP, transformants were secondly patched onto nutrient-complete YPD agar plus G418 (a kanamycin analogue). If cells are G418-resistant and grow on this media, the presence of *KanMX* is indicated.

Following transformation with pRM5, each *KanMX* mutant tested positive for a *URA3*⁺ phenotype when selected on SD-uracil plates, and did not grow on YPD+G418, phenotypically confirming the replacement of *KanMX* with the ϕ C31 LP (*URA3*). These plates are shown in Figure 6.6, including as controls the non-*KanMX* mutant origin *S. cerevisiae* strain (WT) and the untransformed *KanMX* mutants.

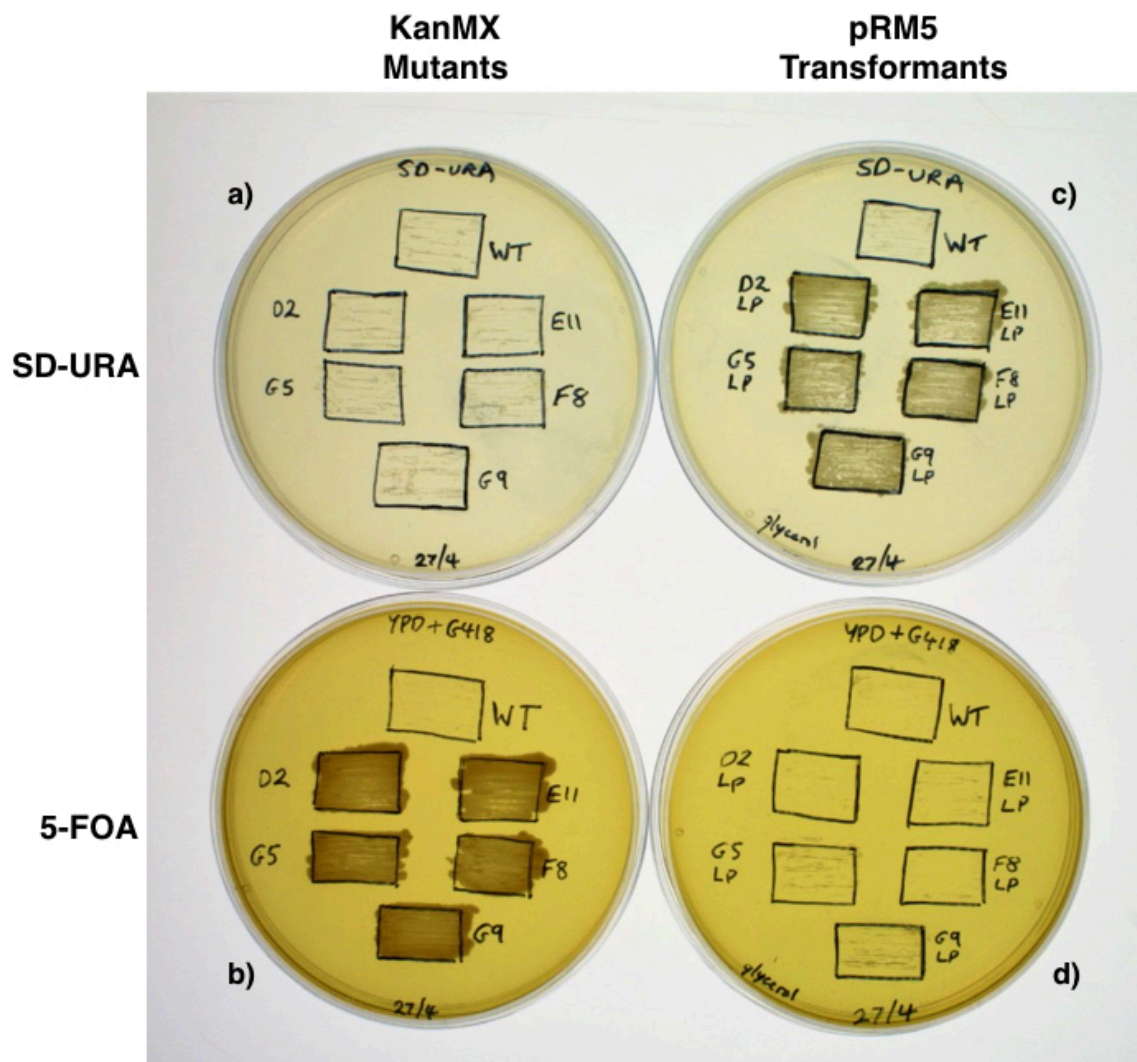


Figure 6.6: Photographs of yeast KanMX mutants transformed with ϕ C31LP homologous recombination plasmid pRM5 (alongside untransformed controls and 'WT' yeast) patched onto confirmation agar plates. Growth on SD-uracil plates indicates presence of the *URA3* marker gene, while growth on YPD + 5-FOA indicates continued presence of the KanMX cassette.

Following the phenotyping of *URA3* LP transformants on selective plates, a further confirmation was done by PCR. Primers were designed to read from outside of the *KanMX* regions up- and down-stream of the LP into the *URA3* gene, as shown in Figure 6.7a. PCRs were performed on a single uracil-positive transformant colony as described in Methods 2.13.1. Both PCRa, reading from an upstream *KanMX* sequence to *URA3*, and PCRb, reading from *URA3* to a downstream *KanMX* sequence, amplified the expected product size in the LP transformants only, with no amplification in the WT background strain or untransformed *KanMX* starter strains. Gel electrophoreses of these PCRs are shown in Figure 6.7.

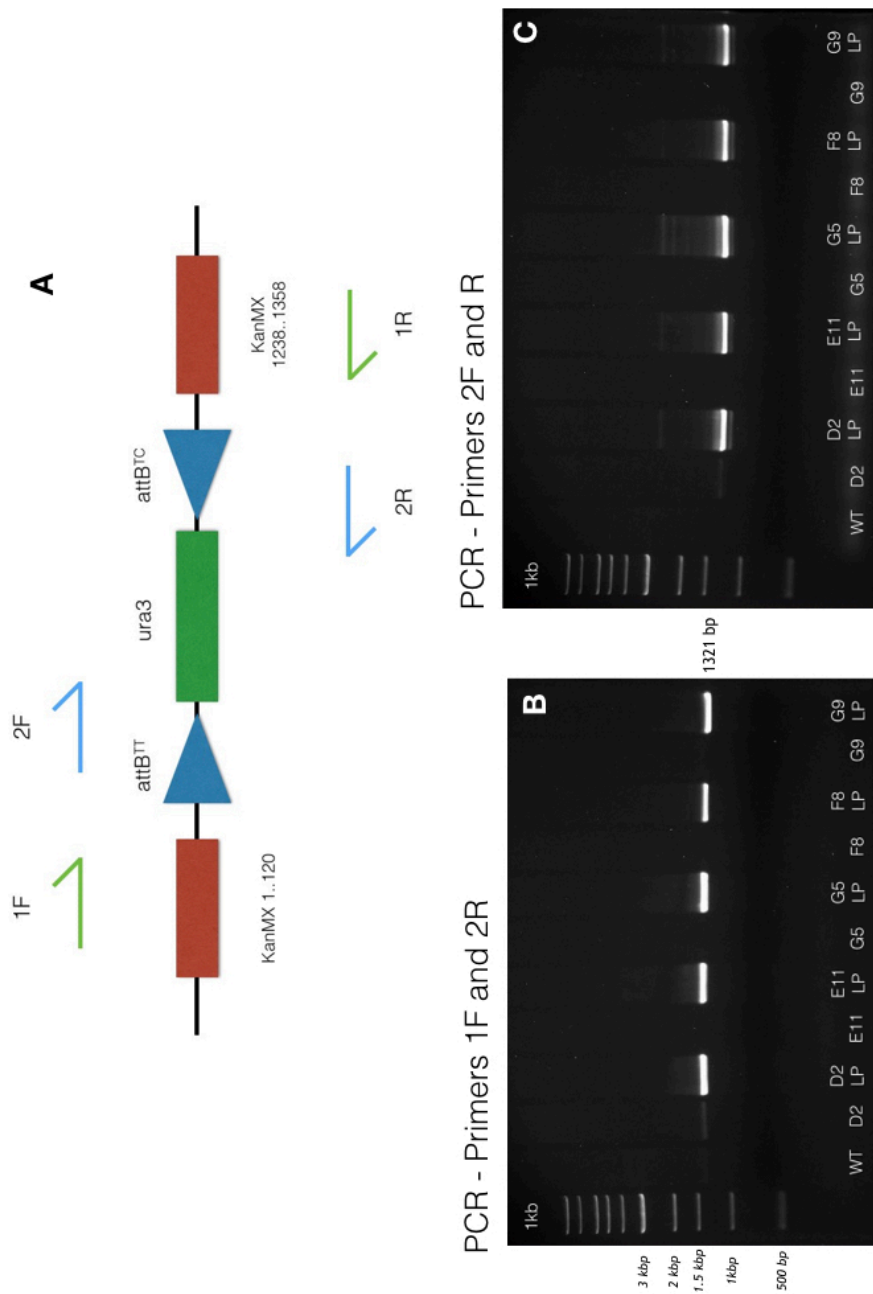


Figure 6.7: Confirmatory PCRs of *KanMX* replacement by ϕ C31 *URA3* LP cassette via homologous recombination in five *KanMX* yeast mutants labelled D2, E11, G5, F8 and G9 (see Figure 6.2); background yeast strain *S. cerevisiae* S288C labelled WT is included in the first lane as a negative control. Uracil-positive pRM5 transformants, labelled D2 LP etc, were tested in two colony PCRs: PCRa with primers 1F and 2R, reading across the *attB^{IT}* flank of the LP (with an expected product of 1321 bp), and PCRb with primers 2F and 1R, reading across the *attB^{TC}* flank of the LP (with an expected product of 1275 bp). Each LP transformant is positive for the expected *URA3*LP PCR band, which is not seen in the untransformed *KanMX* strains (as in the WT control).

While these PCRs appear to confirm presence of the *URA3* LP on the genome of these yeast transformants, it should be noted that due to the *KanMX* primers reading from within the regions of homology used for homologous recombination, there is a small but unlikely chance that the entire cassette could have inserted into a different location on the genome. Further confirmatory PCRs with primers reading from upstream and downstream of the deleted gene region are to be designed. Despite the above caveat, this appears to be a strong indication that the *KanMX* cassette has been successfully replaced by the ϕ C31 LP. Colonies of each confirmed LP mutant were therefore grown in an overnight culture and glycerol stocks were made and stored at -80°C, labelled according to their LP gene deletion reference: F8/YDL229W, D2/YDR138W, G9/YDR007W, E11/YDR297W and G5/YDR539W.

6.3. Discussion

Demonstrating the ϕ C31 integrase LP described in this project in yeast would show its versatility and potential to be used in different industrially-relevant microorganisms, including eukaryotic cells. Much research has been done into the use of serine integrases for genome editing in eukaryotes, however the novelty of this method stems from the fact that the LP system uses two non-identical attachment sites for a single integrase enzyme, rather than two different recombinases as in the RMCE and DICE methods (Turan *et al.*, 2011; Geisinger & Calos, 2014), making this method of genome-based cassette exchange simpler than other systems. With a more complex chromosomal structure and cellular environment, it is likely that position of the integrase LP on the chromosome will have a greater impact on the efficiency of cassette exchange, making it more important to determine locations of high cassette efficiency in the development of a yeast-based LP system. In this work, five ϕ C31 LP mutants were established, with LP locations spread across the length of *S. cerevisiae* S288C chromosome IV (Table 6.1 & Figure 6.2). These ϕ C31LPs, featuring the selectable and counter-selectable *URA3* marker gene, were established via homologous recombination, and confirmed both by

initial phenotypic analysis (for uracil prototrophy and loss of kanamycin resistance; see Figure 6.6) and by PCR confirmation (Figure 6.7).

Once the five ϕ C31 LP mutants were created, some problems were encountered during a wider phenotype of the mutant strains. When cells taken from glycerol stocks of the ϕ C31 LP mutants were tested for auxotrophies, phenotypes were not as expected, bringing into question firstly the correct phenotype background of the *KanMX* strains used here and secondly whether there could have been contamination during creation of the glycerol stocks. At the end of this project, mutants were being further tested to confirm whether they were haploid, as expected, or diploid, which would explain the phenotypes seen during analysis. Depending on these tests, establishment of the correct haploid LP mutants via sporulation would allow subsequent LP cassette exchange experiments to be conducted with a plasmid-borne donor cassette consisting of an antibiotic marker gene flanked by ϕ C31 *attP^{TT}* and *attP^{TC}* sites. The efficiency of these cassette exchange reactions could then be compared between each of the five LP sites on chromosome IV to determine whether genomic location has a significant effect, if any, on integration efficiencies in yeast.

7. Chapter 7: Discussion

7.1. Establishing a ϕ C31 integrase landing pad system in *E. coli*

The overarching aim of this project was to demonstrate a serine integrase landing pad method in *E. coli*, with a view to eventually creating a multi-integrase landing pad system to allow for genomic integration of multiple heterologous genes or whole pathways. Previous work had shown that serine integrases are able to efficiently integrate large lengths of DNA into single sites on the *E. coli* genome with high efficiency (Muroi *et al.*, 2013). The previously developed recombinase-mediated cassette exchange (RMCE) method, in which tyrosine recombinases were used for cassette exchange at a site consisting of two non-identical attachment sites flanking a marker gene (Turan *et al.*, 2011), also formed a basis for developing the serine integrase landing pad system.

In this work, a landing pad cassette was designed for ϕ C31 integrase, consisting of two *attB* sites in a head-to-head orientation (with differing central dinucleotides of TT and TC) surrounding a marker gene for erythromycin resistance. This cassette was inserted into the *E. coli* genome to replace the non-essential *pepA* gene via recombineering. While initial confirmation tests of the recombineering insertion indicated that the LP cassette had been successfully inserted into the *pepA* gene, a 17 bp deletion in the LP's *attB^{TC}* site went undetected during consequent cassette exchange experiments. The lack of cassette exchange detected in the initially-created LPC311a was at first attributed to inactivity of the ϕ C31 integrase enzyme.

Recombination activity assays performed in LPC311a revealed that the ϕ C31 integrase expression plasmid was being broken up and selected against by the cells, suggesting that it was having a toxic effect on cells containing the landing pad. This effect was successfully navigated by switching to an arabinose-inducible ϕ C31 integrase expression plasmid, avoiding the build up

of the enzyme in the cells before it was needed. The system developed here now allows for ϕ C31 integrase expression to be switched on immediately prior to transformation, and resulted in strong activity in *in vivo* recombination assays.

Due to the 17 bp deletion that was found in the ϕ C31 integrase LP *attB*^{TC} site, cassette exchange experiments were performed in *E. coli* cells that contained only on viable *attB*^{TT} site. This meant that the result of donor plasmid transformation, under expression of ϕ C31 integrase, resulted in integration of the complete 4306 bp donor plasmid (pRM3) by *attB*^{TT} x *attP*^{TT} recombination. This considered, the results observed from cassette exchange reactions can be viewed as a demonstration of the efficient integration of 4 kb DNA at the target site. From these experiments, each set of 100 transformant colonies were positive for the three antibiotic resistance markers of the plasmid backbone, donor cassette and LP cassette, indicating that whole plasmid integration at the *pepA* ϕ C31 LP *attB*^{TT} site occurred in all colonies tested. This result is supported by previous genomic integration studies using TG1 integrase which demonstrated integration of entire plasmids into introduced single *attP* or *attB* sites on the *E. coli* genome at high efficiencies (Hirano et al., 2011; Muroi et al., 2013).

It is worth noting that in the time since the start of this project, progress has been made on other cassette exchange methods similar to that of the landing pad system described here. The recombinase-mediated cassette exchange (RMCE) method has been advanced to create a dual RMCE system in which the two attachment sites in the system are for two different tyrosine recombinases, Cre and Flp, rather than variations on the attachment sites of the same recombinase enzyme (Voziyanova et al, 2017). However, the ability of tyrosine recombinases to also catalyse the reverse excision reaction still means that this system is not as efficient as is possible with the more unidirectional serine integrases.

A serine integrase-based genomic cassette exchange method resembling the landing pad system described here has also been developed for gene integrations in mammalian cells. Referred to as dual integrase cassette

exchange (DICE) (Geisinger & Calos, 2014), this system differs in that it uses two integrases for a single landing pad, making use of both a ϕ C31 and a Bxb1 *att* site as part of the same landing pad-like cassette. In this method, a cassette consisting of a ϕ C31 *attP* site and a Bxb1 *attP* site flanking a *kan/neoR* gene and GFP is pre-inserted into the genome of human embryonic stem cells. A donor plasmid featuring the corresponding ϕ C31 *attB* and Bxb1 *attB* sites flanking the target gene and an *mCherry* marker gene is then co-transfected into the cell with expression plasmids for ϕ C31 integrase and Bxb1 integrase, resulting in cassette exchange across the site (Zhu *et al*, 2014; Farruggio *et al*, 2017). This method has shown high specificity and efficiency, however in comparison to the single integrase LP system described in this study, DICE is complicated by the need to transform or transfect two integrase expression vectors alongside the donor plasmid.

Since the start of this project the list of integrases available for use in genome editing has expanded, providing yet further potential options for the integrase LP system. In 2014, Yang *et al* used bioinformatic analysis of prophage genomes from conserved domain databases to identify 34 putative serine integrases and their *attB* and *attP* sites, 11 of which they used to successfully create memory switches in *E. coli* (Yang *et al*, 2014). Another novel serine integrase of potential use to genomic integrations of heterologous DNA that has been characterised in the last year is ϕ Joe, a *Streptomyces* phage-derived integrase isolated by Fogg *et al* (2017) alongside its RDF, gp52. This adds to the list of integrase-RDF pairings available for use in synthetic biology and represents another addition to a possible multi-integrase LP toolkit for synthetic biology.

7.1. Non-cognate activity of integrases and other RDFs

Another primary objective of this work was to study the cross-compatibility of serine integrases and their non-cognate RDFs. In order to develop a multi-integrase LP system, which includes reverse excision cassette exchanges between *attL* and *attR* sites formed following an initial LP cassette exchange, the fidelity of integrases and their RDFs must be explored. To use two

integrases in the same multi-LP cell in this reversible manner, there must be little cross-talk between an integrase and the RDF of the other integrase in the system. Here, ϕ C31 integrase and TG1 integrase were investigated for that cross-reactivity.

Through *in vitro* recombination assays, it was revealed that TG1 RDF gp25 was indeed capable of strongly activating ϕ C31 *attL-attR* recombination, with a 39% recombination of the substrate DNA observed (see Figure 5.2). This means that when establishing a multi-integrase LP system there will be caveats when using these two integrases and their RDFs (for secondary *attL* x *attR* cassette exchange) in the same system. On the other hand, ϕ C31 RDF gp3 was shown to not significantly activate TG1 *attL-attR* recombination, with recombination of the substrate plasmid very close to the base level seen with only the TG1 integrase without addition of any RDF (Figure 5.3).

To support the prospect of a markerless landing pad system with very high cassette exchange efficiencies, the *in vitro* recombination assay data presented here acts to illustrate the strong unidirectionality of the serine integrases. This is particularly true of TG1 integrase, which showed only a very low level of 9% recombination of its *attL* and *attR* sites in the absence of RDF. However, when its RDF gp25 was added, *attL-attR* recombination increased dramatically to 72%. Once this *attL-attR* recombination has taken place, TG1 gp25 then shows very high inhibitory activity of the resulting *attP* and *attB* sites with decrease in *attP-attB* recombination from 77% without gp25 to 8% when the RDF is included, showing the powerful directionality of this integrase. As well as showing promise for the development of an efficient TG1 LP system, this result suggests that TG1 integrase as a prime candidate for development of other synthetic biology technologies due to its strict and responsive directionality control and high rates of recombination.

References

- Baeshen, M.N., Al-Hejin, A.M., Bora, R.S., Ahmed, M.M.M., Ramadan, H.A.I., Saini, K.S., Baeshen, N.A., and Redwan, E.M. (2015) Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. *J Microbiol Biotechnol*, 25, 953-962.
- Bassalo, M.C., Garst, A.D., Halweg-Edwards, A.L., Grau, W.C., Domaille, D.W., Mutalik, V.K., Arkin, A.P., and Gill, R.T. (2016) Rapid and Efficient One-Step Metabolic Pathway Integration in *E. coli*. *ACS Synth Biol*, 5, 561-568.
- Bateman, J.R., Lee, A.M. and Wu, C.T. (2006) Site-specific transformation of *Drosophila* via phi C31 integrase-mediated cassette exchange. *Genetics*, 173, 769-777.
- Bierman, M., Logan, R., Obrien, K., Seno, E.T., Rao, R.N. and Schoner, B.E. (1992) Plasmid Cloning Vectors for the Conjugal Transfer of DNA from *Escherichia-Coli* to *Streptomyces* Spp. *Gene*, 116, 43-49.
- Birnbaum, S. and Bailey, J.E. (1991) Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol Bioeng*, 37, 736-745.
- Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phi C31 integrases. *P Natl Acad Sci USA*, 104, 3312-3317.
- Block, D.H.S., Hussein, R., Liang, L.W. and Lim, H.N. (2012) Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res*, 40, 8979-8992.
- Bryant, J.A., Sellars, L.E., Busby, S.J.W. and Lee, D.J. (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res*, 42, 11383-11392.

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. (1988) The Pmtl Nic-Cloning Vectors .1. Improved Puc Polylinker Regions to Facilitate the Use of Sonicated DNA for Nucleotide Sequencing. *Gene*, 68, 139-149.

Chaverroche, M.K., Ghigo, J.M. and d'Enfert, C. (2000) A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res*, 28.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R. et al. (2012) *Saccharomyces Genome Database: the genomics resource of budding yeast*. *Nucleic Acids Res*, 40, D700-D705.

Colloms, S.D., Merrick, C.A., Olorunniji, F.J., Stark, W.M., Smith, M.C.M., Osbourn, A., Keasling, J.D. and Rosser, S.J. (2014) Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res*, 42.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet*, 2, 769-779.

Das, S., Noe, J.C., Paik, S. and Kitten, T. (2005) An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *J Microbiol Meth*, 63, 89-94.

Demarre, G., Guerout, A.M., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marliere, P. and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncP alpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol*, 156, 245-255.

Dillingham, M.S. and Kowalczykowski, S.C. (2008) RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiol Mol Biol R*, 72, 642-+.

Englaender, J.A., Jones, J.A., Cress, B.F., Kuhlman, T.E., Linhardt, R.J. and Koffas, M.A.G. (2017) Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in E-coli. *Acs Synth Biol*, 6, 710-720.

Farruggio, A.P. and Calos, M.P. (2014) Serine integrase chimeras with activity in E. coli and HeLa cells. *Biol Open*, 3, 895-903.

Geisinger, J.M. and Calos, M.P. (2015) Using Phage Integrases in a Site-Specific Dual Integrase Cassette Exchange Strategy. *Methods Mol Biol*, 1239, 29-38.

Ghosh, P., Kim, A.I. and Hatfull, G.F. (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol Cell*, 12, 1101-1111.

Giaever, G. and Nislow, C. (2014) The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics*, 197, 451-465.

Grindley, N.D., Whiteson, K.L. and Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem*, 75, 567-605.

Groth, A.C., Fish, M., Nusse, R. and Calos, M.P. (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phi C31. *Genetics*, 166, 1775-1782.

Groth, A.C., Olivares, E.C., Thyagarajan, B. and Calos, M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *P Natl Acad Sci USA*, 97, 5995-6000.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-Bad Promoter. *Journal of Bacteriology*, 177, 4121-4130.

Haldimann, A. and Wanner, B.L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol*, 183, 6384-6393.

Hashimoto-gotoh, T. and Sekiguchi, M. (1977) Mutations to Temperature Sensitivity in R-Plasmid Psc101. *Journal of Bacteriology*, 131, 405-412.

Hirano, N., Muroi, T., Kihara, Y., Kobayashi, R., Takahashi, H. and Haruki, M. (2011) Site-specific recombination system based on actinophage TG1 integrase for gene integration into bacterial genomes. *Appl Microbiol Biot*, 89, 1877-1884.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013) CRISPR-assisted editing of bacterial genomes. *Nat Biotechnol*, 31, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.

Katayama, Y., Ito, T., and Hiramatsu, K. (2000) A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother*, 44, 1549-1555.

Khaleel, T., Younger, E., McEwan, A.R., Varghese, A.S. and Smith, M.C.M. (2011) A phage protein that binds phi C31 integrase to switch its directionality. *Molecular Microbiology*, 80, 1450-1463.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev*, 58, 401-465.

Landy, A. (1989) Dynamic, Structural, and Regulatory Aspects of Lambda-Site-Specific Recombination. *Annual Review of Biochemistry*, 58, 913-949.

Lewis, J.A., and Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res*, 29, 2205-2216.

Lin-Chao, S. and Bremer, H. (1986) Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. *Mol Gen Genet*, 203, 143-149.

Loessner, M.J., Inman, R.B., Lauer, P., and Calendar, R. (2000) Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol Microbiol*, 35, 324-340.

Luo, M.L., Leenay, R.T., and Beisel, C.L. (2016) Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol Bioeng*, 113, 930-943.

Martinez-Morales, F., Borges, A.C., Martinez, K., Shanmugam, K.T. and Ingram, L.O. (1999) Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction. *Journal of Bacteriology*, 181, 7143-7148.

Morita, K., Yamamoto, T., Fusada, N., Komatsu, M., Ikeda, H., Hirano, N. and Takahashi, H. (2009) The site-specific recombination system of actinophage TG1. *Fems Microbiol Lett*, 297, 234-240.

Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. and Church, G.M. (2012) Improving Lambda Red Genome Engineering in *Escherichia coli* via Rational Removal of Endogenous Nucleases. *Plos One*, 7.

Mouw, K.W., Rowland, S.-J., Gajjar, M.M., Boocock, M.R., Stark, W.M., and Rice, P.A. (2008) Architecture of a serine recombinase-DNA regulatory complex. *Mol Cell*, 30, 145-155.

Müller-Hill, B., Crapo, L., and Gilbert, W. (1968) Mutants that make more lac repressor. *Proc Natl Acad Sci U.S.A.*, 59, 1259-1264.

Muroi, T., Kokuzawa, T., Kihara, Y., Kobayashi, R., Hirano, N., Takahashi, H. and Haruki, M. (2013) TG1 integrase-based system for site-specific gene integration into bacterial genomes. *Appl Microbiol Biot*, 97, 4039-4048.

Nair, J., Rouse, D.A., Bai, G.H. and Morris, S.L. (1993) The Rpsl Gene and Streptomycin Resistance in Single and Multiple Drug-Resistant Strains of *Mycobacterium-Tuberculosis*. *Molecular Microbiology*, 10, 521-527.

Olivares, E.C., Hollis, R.P. and Calos, M.P. (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene*, 278, 167-176.

Peng, R., Lin, G., and Li, J. (2016) Potential pitfalls of CRISPR/Cas9-mediated genome editing. *FEBS J*, 283, 1218-1231.

Pyne, M.E., Moo-Young, M., Chung, D.A., and Chou, C.P. (2015) Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in *Escherichia coli*. *Appl Environ Microbiol*, 81, 5103-5114.

Reijns, M., Lu, Y.J., Leach, S. and Colloms, S.D. (2005) Mutagenesis of *PepA* suggests a new model for the Xer/cer synaptic complex. *Molecular Microbiology*, 57, 927-941.

Ritacco, C.J., Kamtekar, S., Wang, J.M. and Steitz, T.A. (2013) Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase. *Nucleic Acids Res*, 41, 2673-2682.

Ronda, C., Maury, J., Jakočiunas, T., Jacobsen, S.A.B., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K., and Nielsen, A.T. (2015) CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact*, 14, 97.

Rosano, G.L. and Ceccarelli, E.A. (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*, 5.

Rowley, P.A., Smith, M.C.A., Younger, E. and Smith, M.C.M. (2008) A motif in the C-terminal domain of phi C31 integrase controls the directionality of recombination. *Nucleic Acids Res*, 36, 3879-3891.

Russell, J.P., Chang, D.W., Tretiakova, A. and Padidam, M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques*, 40, 460-+.

Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G.D. (2013) Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res*, 41, 8341-8356.

Rutherford, K. and Van Duyne, G.D. (2014) The ins and outs of serine integrase site-specific recombination. *Curr Opin Struc Biol*, 24, 125-131.

Sahdev, S., Khattar, S.K., and Saini, K.S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem*, 307, 249-264.

Sauer, B. (1998) Inducible gene targeting in mice using the Cre/lox system. *Methods*, 14, 381-392.

Schlake, T. and Bode, J. (1994) Use of Mutated FLP Recognition Target (FRT) Sites for the Exchange of Expression Cassettes at Defined Chromosomal Loci. *Biochemistry-US*, 33, 12746-12751.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol*, 189, 8746-8749.

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. and Court, D.L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc*, 4, 206-223.

- Shiloach, J. and Fass, R. (2005) Growing E-coli to high cell density - A historical perspective on method development. *Biotechnol Adv*, 23, 345-357.
- Singh, S., Rockenbach, K., Dedrick, R.M., VanDemark, A.P. and Hatfull, G.F. (2014) Cross-talk between Diverse Serine Integrases. *J Mol Biol*, 426, 318-331.
- Smith, M.C. and Thorpe, H.M. (2002) Diversity in the serine recombinases. *Mol Microbiol*, 44, 299-307.
- Smith, M.C.M. (2015) Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiol Spectr*, 3.
- Storici, F., Durham, C.L., Gordenin, D.A., and Resnick, M.A. (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci, U.S.A.*, 100, 14994-14999.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S. and Calos, M.P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phi C31 integrase. *Mol Cell Biol*, 21, 3926-3934.
- Turan, S., Galla, M., Ernst, E., Qiao, J.H., Voelkel, C., Schiedlmeier, B., Zehe, C. and Bode, J. (2011) Recombinase-Mediated Cassette Exchange (RMCE): Traditional Concepts and Current Challenges. *J Mol Biol*, 407, 193-221.
- Urasaki, A., Sekine, Y. and Ohtsubo, E. (2002) Transposition of cyanobacterium insertion element ISY100 in *Escherichia coli*. *Journal of Bacteriology*, 184, 5104-5112.
- Wang, H., and Mullany, P. (2000) The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397. *J Bacteriol*, 182, 6577-6583.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460, 894-U133.

Wang, H., La Russa, M., and Qi, L.S. (2016) CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem*, 85, 227-264.

Xu, Z., Lee, N.C., Dafhnis-Calas, F., Malla, S., Smith, M.C. and Brown, W.R. (2008) Site-specific recombination in *Schizosaccharomyces pombe* and systematic assembly of a 400kb transgene array in mammalian cells using the integrase of *Streptomyces* phage phiBT1. *Nucleic Acids Res*, 36, e9.

Yuan, P., Gupta, K., and Van Duyne, G.D. (2008) Tetrameric structure of a serine integrase catalytic domain. *Structure*, 16, 1275-1286.

Zerbini, F., Zanella, I., Fraccascia, D., König, E., Irene, C., Frattini, L.F., Tomasi, M., Fantappiè, L., Ganfini, L., Caproni, E., et al. (2017). Large scale validation of an efficient CRISPR/Cas-based multi gene editing protocol in *Escherichia coli*. *Microb Cell Fact*, 16, 68.

Zhang, L., Zhu, B.Y., Dai, R.X., Zhao, G.P. and Ding, X.M. (2013) Control of Directionality in *Streptomyces* Phage phi BT1 Integrase-Mediated Site-Specific Recombination. *Plos One*, 8.